

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:59:05 ; Search time 27 seconds

(without alignments)  
1496.535 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 776

Sequence: 1 MTGIAAAGFFSNTCRFGCG.....KALETDSVSGVSKQKQKL 776

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647	83.4	739	1 JUAZ_HUMAN	Q15022 homo sapien
2	272	35.1	741	1 JUAZ_MOUSE	Q80u70 mus musculus
3	128	16.5	243	1 JAZ1_HUMAN	Q86vz6 homo sapien
4	128	16.5	243	1 JAZ1_MOUSE	Q80zq5 mus musculus
5	11	1.4	900	1 SU12_DROME	Q9njg9 drosophila
6	8	1.0	247	1 Y124_RPVAC	P41679 autographa
7	8	1.0	662	1 GGT4_MOUSE	Q99jp7 mus musculus
8	8	1.0	662	1 GGT4_RAT	Q99mz4 rattus norv
9	8	1.0	666	1 GGT4_HUMAN	Q9u114 homo sapien
10	8	1.0	669	1 COTE_HUMAN	P81408 homo sapien
11	8	1.0	1175	1 P1B4_HUMAN	Q15147 homo sapien
12	7	0.9	46	1 THN2_VISAL	P32880 viscum albu
13	7	0.9	46	1 THN_PHOTO	P01539 phoradendro
14	7	0.9	105	1 SUG2_CITFR	Q69279 citrobacter
15	7	0.9	139	1 ACSB_BRAJA	Q69159 bradyrhizob
16	7	0.9	152	1 ASNC_ECOLI	P03809 escherichia
17	7	0.9	153	1 E417_ADE02	P03238 human adeno
18	7	0.9	154	1 NHP2_SCHPO	Q9p7h0 schizosach
19	7	0.9	159	1 MUTX_STRMO	P95781 streptococc
20	7	0.9	205	1 RUVA_BACCA	Q81l98 bacillus ce
21	7	0.9	205	1 RUVA_BACCA	Q81w33 bacillus ce
22	7	0.9	234	1 POLN_RHDV3	P27411 rabbit hemo
23	7	0.9	243	1 NUKS_RAT	Q9epj0 rattus norv
24	7	0.9	269	1 H1S9_LACLA	Q02150 lactococcus
25	7	0.9	269	1 TH1D_STRCO	Q9zbr6 streptomyce
26	7	0.9	274	1 ABCD_HUMAN	Q9nse7 homo sapien
27	7	0.9	293	1 ERA_MYCPU	Q98q11 mycoplasma
28	7	0.9	294	1 E434_ADE02	P03239 human adeno
29	7	0.9	319	1 YC39_ODOSI	P49534 odontella s
30	7	0.9	324	1 ANXA_HUMAN	Q9uj72 homo sapien
31	7	0.9	329	1 YG54_ARCFU	Q28619 archaeoglob
32	7	0.9	368	1 SH31_HUMAN	Q99961 homo sapien
33	7	0.9	368	1 SH31_MOUSE	Q62419 mus musculus

34	7	0.9	377	1 GSB_ARATH	P49177 arabidopsis
35	7	0.9	377	1 PEL_XANCM	Q56806 xanthomonas
36	7	0.9	378	1 PEL_PSESL	P72242 pseudomonas
37	7	0.9	380	1 AROF_CHLPLN	Q926m3 chlamydia p
38	7	0.9	380	1 PEL_PSEPL	Q59671 pseudomonas
39	7	0.9	380	1 PEL_PSEVI	Q60140 pseudomonas
40	7	0.9	391	1 EFTU_AGRTS	Q8ue16 agrobacteri
41	7	0.9	391	1 EFTU_AGRTU	P75022 agrobacteri
42	7	0.9	391	1 EFTU_BRUME	Q8yh04 brucella me
43	7	0.9	391	1 EFTU_RHIME	Q925y6 rhizobium m
44	7	0.9	395	1 EFTU_CYTLY	P42474 cytophaga l
45	7	0.9	395	1 EFTU_PSEPK	Q98qp8 pseudomonas

## ALIGNMENTS

RESULT 1  
JUAZ\_HUMAN  
ID JUAZ\_HUMAN STANDARD; PRT; 739 AA.  
AC Q15022; Q96BD9;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Polycarb protein SUZ12 (Suppressor of zeste 12 protein homolog)  
DE (Joined to JAZF1 protein) (Chromatin precipitated E2F target 9  
protein) (ChET 9 protein).  
GN SUZ12 OR JUAZ1 OR CHET9 OR KIAA0160.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;  
MEDLINE=96127530; PubMed=8590280;  
RT "Prediction of the coding sequences of unidentified human genes. IV.  
analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 2:167-174(1995).  
RX SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
MEDLINE=22388257; PubMed=12477932;  
RA Klausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
Gnenerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RX [3]  
DE SEQUENCE OF 102-113; 343-360; 474-483 AND 509-516, IDENTIFICATION IN A  
PRC2 COMPLEX WITH E2F; E2H2; RBBP4 AND RBBP7, AND METHYLTRANSFERASE  
ACTIVITY OF THE COMPLEX.  
DE MEDLINE=22300550; PubMed=12351676;  
RX Cao R., Wang L., Wang H., Xia L., Erdjument-Bromage H., Tempst P.,  
Jones R.S., Zhang Y.;  
RT "Role of histone H3 lysine 27 methylation in Polycomb-group

silencing.";  
 RL Science 298:1039-1043 (2002).  
 [4]  
 RP DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH JAZF1.  
 RX MEDLINE=21265470; PubMed=11371647;  
 RA Krontz J.I., Soreng A.L., Nucci M., Kuo F.C., Pauwels P.,  
 Van Den Berghe H., Clin P.D., Fletcher J.A., Sklar J.;  
 RT "Frequent fusion of the JAZF1 and JUAZ1 genes in endometrial stromal  
 tumors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353 (2001).  
 [5]  
 RP INDUCTION.  
 RX MEDLINE=21448978; PubMed=11564866;  
 RA Weinmann A.S., Bartley S.M., Zhang T., Zhang M.Q., Farnham P.J.;  
 RT "Use of chromatin immunoprecipitation to clone novel E2F target  
 promoters.";  
 RL Mol. Cell. Biol. 21:6820-6832 (2001).  
 [6]  
 RP IDENTIFICATION IN A PRC2 COMPLEX WITH EED; EZH2; RBBP4 AND RBBP7, AND  
 METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
 RX MEDLINE=22323099; PubMed=12435631;  
 RA Kuzmichev A., Nishio K., Erdjument-Bromage H., Tempst P.,  
 Reinberg D.;  
 RT "Histone methyltransferase activity associated with a human  
 multiprotein complex containing the Enhancer of Zeste protein.";  
 RL Genes Dev. 16:2893-2905 (2002)  
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by  
 forming multiprotein complexes, which are required to maintain the  
 transcriptionally repressive state of homeotic genes throughout  
 development. PcG proteins are not required to initiate repression,  
 but to maintain it during later stages of development. They  
 probably act via the methylation of histones, rendering chromatin  
 heritably changed in its expressibility. Component of the PRC2  
 complex, which methylates Lys-9 and Lys-27 residues of histone H3.  
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,  
 SUZ12/JAZF1, RBBP7. This complex is distinct from the  
 PRC1 complex, which contains many other PcG proteins like BMI1,  
 CBX4, CBX6, PHC2, PHC3, SGMH1, RING1 and RNF2.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- INDUCTION: Induced by E2F1 transcription factor.  
 CC -!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal  
 rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21)  
 chromosomal translocation involving JAZF1 and SUZ12 generates the  
 JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1  
 and the C-terminus part of SUZ12. The translocation is frequently  
 found in all cases of endometrial stromal tumors, except in  
 endometrial stromal sarcomas where it is rarer.  
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(2)12) family.  
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; D63881; BAA09931.1; ALT INIT.  
 DR EMBL; BC015704; AAH15704.1; -.  
 DR MIM; 606245; -.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
 KW Transcription regulation; Repressor; Chromatin regulator;  
 KW Nuclear protein; Metal-binding; Zinc; Zinc-finger;  
 KW Chromosomal translocation; Proto-oncogene.  
 FT ZN FING 448 471 C2H2-TYPE.  
 FT DOMAIN 563 639 VEFs-BOX.  
 FT DOMAIN 7 50 GLY-RICH.  
 FT DOMAIN 51 59 POLY-SER.  
 FT DOMAIN 60 67 POLY-ALA.  
 FT SITE 93 94 BREAKPOINT FOR TRANSLOCATION TO FORM

FT JAZF1-SUZ12 ONCOGENE.  
 SQ SEQUENCE 739 AA; 83053 MW; A8830BBD031DF656 CRC64;  
 Query Match 83.4%; Score 647; DB 1; Length 739;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 130 PTQIVRFLRLNLIAPIFLHRTLYNTHSRNSTNKRKTFKVDMLSKVKMKGEQESH 189  
 DB 93 PTQIVRFLRLNLIAPIFLHRTLYNTHSRNSTNKRKTFKVDMLSKVKMKGEQESH 152  
 QY 190 LSAHLQLTGTFPHNDKPSNSENEQNSVTLVLLVVKVCHKKRDXSCPIRQVPTGKK 249  
 DB 153 LSAHLQLTGTFPHNDKPSNSENEQNSVTLVLLVVKVCHKKRDXSCPIRQVPTGKK 212  
 QY 250 VPLIPDLNQTGPNFSLAVSNSEPEPNSHMKVSYLLFVTRPGRREFNGMINGETNE 309  
 DB 213 VPLIPDLNQTGPNFSLAVSNSEPEPNSHMKVSYLLFVTRPGRREFNGMINGETNE 272  
 QY 310 NIDVNEELPARKNRDEGEKTFVAQMTVFDKRRLOLLDGEYEVAMQEMEECPISKRA 369  
 DB 273 NIDVNEELPARKNRDEGEKTFVAQMTVFDKRRLOLLDGEYEVAMQEMEECPISKRA 332  
 QY 370 TWETILDGRLPPPTFFSQPTLQTLRWGTGNDKSTAPIAKPLATRNSSLHQNKP 429  
 DB 333 TWETILDGRLPPPTFFSQPTLQTLRWGTGNDKSTAPIAKPLATRNSSLHQNKP 392  
 QY 430 SVKPTQTIKVESLTDTLOTRKEKTPNENROKLRIFQFLYNNTRQOTTEARDLHCPW 489  
 DB 393 SVKPTQTIKVESLTDTLOTRKEKTPNENROKLRIFQFLYNNTRQOTTEARDLHCPW 452  
 QY 490 CTLNCRKLYSLKHLKCHSRFIFNVYHPKGRDVSINECYDGSYAGNPQDIHQPGF 549  
 DB 453 CTLNCRKLYSLKHLKCHSRFIFNVYHPKGRDVSINECYDGSYAGNPQDIHQPGF 512  
 QY 550 AFSRNGPVKRTPIHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHNRLYFHSDFCL 609  
 DB 513 AFSRNGPVKRTPIHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHNRLYFHSDFCL 572  
 QY 610 PIRPQMEVDSDEKDPWLREKTIQTEEFSDVNEGEKVMKLNHVMKHGFTADNQ 669  
 DB 573 PIRPQMEVDSDEKDPWLREKTIQTEEFSDVNEGEKVMKLNHVMKHGFTADNQ 632  
 QY 670 NEACMLFVENVYQKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLRMQQKLEGE 729  
 DB 633 NEACMLFVENVYQKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLRMQQKLEGE 692  
 QY 730 SASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
 DB 693 SASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 739  
 RESULT 2  
 JUAZ\_MOUSE  
 ID JUAZ\_MOUSE STANDARD; PRT; 741 AA.  
 AC Q80U70; Q80U10; Q99L07;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Polycomb protein Suz12 (Suppressor of zeste 12 protein homolog).  
 GN SUZ12 OR KIAA0160 OR D11ERTD530E.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RX MEDLINE=22579291; PubMed=12693553;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,  
 Nakajima D., Nagase T., Ohara O., Koga H.;  
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:



RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous  
 RT cDNAs identified by screening of terminal sequences of cDNA clones  
 RT randomly sampled from size-fractionated libraries."  
 RL DNA Res. 10:35-48(2003).  
 RN [2]  
 RP SEQUENCE OF 365-741 FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by  
 CC forming multiprotein complexes, which are required to maintain the  
 CC transcriptionally repressive state of homeotic genes throughout  
 CC development. PcG proteins are not required to initiate repression,  
 CC but to maintain it during later stages of development. They  
 CC probably act via the methylation of histones, rendering chromatin  
 CC heritably changed in its expressibility. Component of the PRC2  
 CC complex, which methylates Lys-9 and Lys-27 residues of histone H3  
 CC (By similarity).  
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,  
 CC SUZ12/JAZ1, RBBP4 and RBBP7. This complex is distinct from the  
 CC PRC1 complex, which contains many other PcG proteins like EMI1,  
 CC CBX4, CBX6, PRC2, PRC3, SCMH1, RING1 and RNF2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(2)12) family.  
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to an  
 CC erroneous RNA joining and initiations.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AK122213; BAC65495.1; ALT\_SEQ.  
 DR EMBL; BC003922; AAH03922.1; -.  
 DR EMBL; BC051099; AAH51099.1; -.  
 DR MGD; MGI:1261758; D11Ertd530e.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
 KW Transcription regulation; Repressor; Chromatin regulator;  
 KW Nuclear protein; Metal-binding; Zinc; Zinc-finger.  
 FT ZN FING 450 473  
 FT DOMAIN 565 641 VEFs-BOX.  
 FT DOMAIN 7 50 GLY-RICH.  
 FT DOMAIN 29 63 AUA-RICH.  
 SQ SEQUENCE 741 AA; 83025 MW; 4ACB6E52D24FAEF2 CRC64;  
 Query Match 35.1%; Score 272; DB 1; Length 741;  
 Best Local Similarity 100.0%; Pred. No. 7.5e-276;  
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 RQKLRIYQFLYNNNTQQTEARDLHCPWCTLCNCRKLYSLKHLKLCRSRFFIYVYHP 519  
 DB 425 RQKLRIYQFLYNNNTQQTEARDLHCPWCTLCNCRKLYSLKHLKLCRSRFFIYVYHP 484  
 QY 520 KGARIDVINECYDGSVAGNPQDIHQPGFAPGRNGPVKPTPTTHILVCPKPKTKASMS 579  
 DB 485 KGARIDVINECYDGSVAGNPQDIHQPGFAPGRNGPVKPTPTTHILVCPKPKTKASMS 544  
 QY 580 FLESEDEGEVQOYRTSYSGHNRLYFHSHTCLPLRPOEMEVSEDEKDPENLREKTIIOEE 639  
 DB 545 FLESEDEGEVQOYRTSYSGHNRLYFHSHTCLPLRPOEMEVSEDEKDPENLREKTIIOEE 604  
 QY 640 FSDVNEGEKVEYKWLWNJHVMKGFADNQNHACMLFVENYQKIKKNCRNFMHLVS 699  
 DB 605 FSDVNEGEKVEYKWLWNJHVMKGFADNQNHACMLFVENYQKIKKNCRNFMHLVS 664  
 QY 700 MHDNPLNISIDKAVTKLREMOQKLEKGESA 731  
 DB 665 MHDNPLNISIDKAVTKLREMOQKLEKGESA 696  
 RESULT 3  
 JAZ1 HUMAN  
 ID JAZ1 HUMAN STANDARD; PRT; 243 AA.  
 AC Q86VZ6; G8N3L7;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Juxtaposed with another zinc finger protein 1.  
 GN JAZF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain, and Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Amvgdala;  
 RL Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH SUZ12.  
 RX MEDLINE=21265470; PubMed=11371647;  
 RA Koonz J.I., Soreng A.L., Nucci M., Kuo F.C., Pauwels P.,  
 RA van Den Berghe H., Cin P.D., Fletcher J.A., Sklar J.;  
 RT "Frequent fusion of the JAZF1 and JAZ1 genes in endometrial stromal  
 RT tumors."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353(2001).  
 CC -!- FUNCTION: Potential transcription factor.



This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AX043767; BAC31649.1; -  
EMBL; AX050474; BAC34275.1; -  
EMBL; BC048577; AAH49577.1; -  
InterPro: IPR007087; Znf\_C2H2.  
Pfam: PF00096; zf-C2H2; 3.  
SMART; SW00355; Znf\_C2H2; 3.  
PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2.  
PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
KW Transcription regulation; Nuclear protein; Metal-binding; Zinc;  
Zinc-finger; Alternative splicing  
FT ZN\_FING 12 37 C2H2-TYPE 1.  
FT ZN\_FING 173 198 C2H2-TYPE 2.  
FT ZN\_FING 208 230 C2H2-TYPE 3 (DEGENERATE).  
FT VARSPPLIC 64 64 Missing (in isoform 2).  
FT CONFLICT 211 211 R -> L (in REF. 1; BAC34275).  
SQ SEQUENCE 243 AA; 27097 MW; 878A6EE3D82C0348 CRC64;

Query Match 16.5%; Score 128; DB 1; Length 243;  
Best Local Similarity 100.0%; Pred. No. 1.2e-125;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFSTCTCGGGLHPPTLADLIEHNHDTDRVLEKELQOPTVALSY 60  
DB 1 MTGIAAASFSTCTCGGGLHPPTLADLIEHNHDTDRVLEKELQOPTVALSY 60  
QY 61 INRFMTDAARRQESLKKIQKLSLTSSSVSGNVSTPPRHSGSLTPPTPTPSS 120  
DB 61 INRFMTDAARRQESLKKIQKLSLTSSSVSGNVSTPPRHSGSLTPPTPTPSS 120  
QY 121 SFRSSTPT 128  
DB 121 SFRSSTPT 128

## RESULT 5

SU12\_DROME  
ID SU12\_DROME STANDARD; PRT; 900 AA.  
AC Q9NIG9; Q879D8; Q9VW55;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Polycomb protein Su(z)12 (Suppressor 12 of zeste protein).  
GN SU(z)12 OR CG8013  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT  
RP SU(z)12-2.  
RX MEDLINE=21430867; PubMed=11546753;  
RA Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,  
RA Rasmuson-Lestander A., Mueller J.;  
RT "Su(z)12, a novel Drosophila Polycomb group gene that is conserved in  
RT vertebrates and plants.";  
RL Development 128:3371-3379 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [3]  
RP REVISIONS, AND ALTERNATIVE SPLICING.  
RX MEDLINE=22426063; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX STRAIN=Berkley; TISSUE=Embryo;  
MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A Drosophila full-length cDNA resource.";  
RN Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).  
RN [5]  
RP IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAP1 AND ESC,  
RP AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
RX MEDLINE=22296673; PubMed=12408863;  
RA Czermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V.;  
RT "Drosophila enhancer of Zeste/ESC complexes have a histone H3  
RT methyltransferase activity that marks chromosomal Polycomb sites.";  
RL Cell 111:185-196 (2002).  
RN [6]  
RP IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAP1 AND ESC, AND  
RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
RX MEDLINE=22296674; PubMed=12408864;  
RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,  
RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;  
RT "Histone methyltransferase activity of a Drosophila Polycomb group  
RT repressor complex.";



OX NCBI\_TaxID=10090;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikiado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schombach C., Gofjohori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Flecher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Kung'u D.R., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Mulcais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilning L.G., Wynshaw-Boris A., Yanagisawa M., Yang J., Yang L.,  
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.,  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast cancer;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP SEQUENCE OF 552-662 FROM N.A.  
 RC STRAIN=ISS, and ILS;  
 RX MEDLINE=21363810; PubMed=11471062;  
 RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,  
 RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;  
 RT "High-throughput sequence identification of gene coding variants  
 RT within alcohol-related QTLs";  
 RL Mamm. Genome 12:657-663 (2001).  
 CC -!- FUNCTION: Cleaves glutathione conjugates (By similarity).  
 CC peptide + 5-L-glutamyl-amino acid.  
 CC -!- PATHWAY: Glutathione metabolism.  
 CC -!- SUBUNIT: Heterodimer composed of the light and heavy chains. The  
 CC active site is located in the light chain. Interacts with CTL20

CC (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).  
 CC -!- SIMILARITY: Belongs to the gamma-glutamyltransferase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AK032051; BAC27672.1; -  
 CC EMBL; BC005772; AAH05772.1; -  
 CC EMBL; AF332053; AAK56082.1; -  
 CC EMBL; AF332054; AAK56083.1; -  
 CC MGD; MGI:1913385; Ggt13.  
 CC InterPro; IPR000101; Peptidase\_T3.  
 CC Pfam; PF01019; G\_glu\_transferase\_1.  
 CC PRINTS; PR01210; GGTTRANSFERASE.  
 CC PROSITE; PS00462; G\_GLU\_TRANSFERPEPTIDASE; FALSE NEG.  
 CC Glutathione biosynthesis; Transferase; Acyltransferase; Signal-anchor;  
 CC Transmembrane; Zymogen; Glycoprotein.  
 CC CHAIN 1 472  
 CC GAMMA-GLUTAMYLTRANSFERASE 4 HEAVY  
 CC CHAIN (BY SIMILARITY).  
 CC GAMMA-GLUTAMYLTRANSFERASE 4 LIGHT  
 CC CHAIN (BY SIMILARITY).  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC DOMAIN 128 662  
 CC EXTRACELLULAR (POTENTIAL).  
 CC FT CARBOHYD 198 198  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 267 267  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 283 283  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 330 330  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 353 353  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 394 394  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 519 519  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 523 523  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CARBOHYD 586 586  
 CC N-LINKED (GLCNAC. ) (POTENTIAL).  
 CC FT CONFLICT 148 148  
 CC H -> S (IN REF. 1).  
 CC SQ SEQUENCE 662 AA; 70301 MW; BA7F32512CC4F381 CRC64;  
 Query Match 1.0%; Score 8; DB 1; Length 662;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 172 DMLSKVE 179  
 Db 436 DMLSKVE 443  
 RESULT 8  
 GGT4 RAT  
 ID GGT4 RAT STANDARD; PRT; 662 AA.  
 AC Q99M24;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DE Gamma-glutamyltransferase 4 precursor (EC 2.3.2.2) (Gamma-  
 DE glutamyltransferase 4) (Gamma-glutamyltransferase-like 3).  
 GN GGT13.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 3-662 FROM N.A.  
 RA Yamaguchi T., Araki K., Nawa H.;  
 RT "Rattus norvegicus gamma-glutamyltransferase homolog mRNA complete  
 RT code";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Cleaves glutathione conjugates (By similarity).

CC -!- CATALYTIC ACTIVITY: (5-L-glutamyl)-peptide + an amino acid =  
 CC peptide + 5-L-glutamyl-amino acid.  
 CC -!- PATHWAY: Glutathione metabolism.  
 CC -!- SUBUNIT: Heterodimer composed of the light and heavy chains. The  
 CC active site is located in the light chain. Interacts with CT120  
 CC (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).  
 CC -!- SIMILARITY: Belongs to the gamma-glutamyltransferase family.  
 CC -!- CAUTION: The N-terminus was extended based on the genomic  
 CC sequence, in analogy to ortholog sequences.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: AF244973; AAK27971.1; ALT INIT.  
 CC InterPro: IPR000101; Peptidase\_F3.  
 CC Pfam: PF01019; G-glu\_transferpep; 1.  
 CC PRINTS: PR01210; GGTRANSFERASE.  
 CC PROSITE: PS00462; G\_GLU\_TRANSFERPEPTIDASE; FALSE NEG.  
 CC KX Glutathione biosynthesis; Transferase; Acyltransferase; Signal-anchor;  
 CC Transmembrane; Zymogen; Glycoprotein.  
 CC FT CHAIN 1 472  
 CC GAMMA-GLUTAMYLTRANSFERASE 4 HEAVY  
 CC FT CHAIN 473 662  
 CC GAMMA-GLUTAMYLTRANSFERASE 4 LIGHT  
 CC FT CHAIN (BY SIMILARITY).  
 CC FT CHAIN (BY SIMILARITY).  
 CC FT DOMAIN 1 106  
 CC CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 107 127  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC FT DOMAIN 128 562  
 CC EXTRACELLULAR (POTENTIAL).  
 CC FT CARBOHYD 198 198  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 267 267  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 283 283  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 330 330  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 353 353  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 394 394  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 519 519  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 523 523  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 586 586  
 CC N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC SQ SEQUENCE 662 AA; 70382 MW; F98B7643700BB03A CRC64;  
 CC  
 CC Query Match 1.0%; Score 8; DB 1; Length 662;  
 CC Best Local Similarity 100.0%; Pred. No. 10;  
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC OY 172 DDMLSKVE 179  
 CC |||||  
 CC Db 436 DDMLSKVE 443  
 CC  
 CC RESULT 9  
 CC GGT4 HUMAN  
 CC ID GGT4 HUMAN STANDARD; PRT; 666 AA.  
 CC AC Q9JUL4; Q8NF66; Q9BYP5; Q9BYP6;  
 CC DT 10-OCT-2003 (Rel. 42, Created)  
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE Gamma-glutamyltransferase 4 precursor (EC 2.3.2.2) (Gamma-  
 CC DE glutamyltransferase 4) (Gamma-glutamyltransferase-like 3)  
 CC GN GGT4 OR GGT4.5  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC [1]  
 CC RN SEQUENCE FROM N.A. (ISOFORM 5), AND INTERACTION WITH CT120.  
 CC RX MEDLINE=22230983; PubMed=12270127;  
 CC RA He X., Di Y., Li J., Xie Y., Tang Y., Zhang F., Wei L., Zhang Y.,

RA Qin W., Huo K., Li Y., Wan D., Gu J.;  
 RT "Molecular cloning and characterization of CT120, a novel  
 RT membrane-associated gene involved in amino acid transport and  
 RT glutathione metabolism.";  
 RL Biochem. Biophys. Res. Commun. 297:528-536(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE=Spleen;  
 RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,  
 RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,  
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 6).  
 RC TISSUE=Spleen;  
 RX MEDLINE=22579292; PubMed=12693554;  
 RA Jikuya H., Takano J., Kikuno R., Hirose M., Nagase T., Nomura N.,  
 RA Ohara O.;  
 RT "Characterization of long cDNA clones from human adult spleen. II. The  
 RT complete sequences of 81 cDNA clones.";  
 RL DNA Res. 10:49-57(2003).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,  
 RA Bailey J., Barlow K.P., Beare K.N., Beare D.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark A.G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Levaeslath M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McKerrison T.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 CC -!- FUNCTION: Cleaves glutathione conjugates (By similarity).  
 CC -!- CATALYTIC ACTIVITY: (5-L-glutamyl)-peptide + an amino acid =  
 CC peptide + 5-L-glutamyl-amino acid.  
 CC -!- PATHWAY: Glutathione metabolism.  
 CC -!- SUBUNIT: Heterodimer composed of the light and heavy chains. The  
 CC active site is located in the light chain (By similarity).  
 CC Isoform 5 interacts with CT120.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Name=1; Synonyms=a;  
 CC Note=Gene prediction confirmed by EST data;  
 CC Name=2; Synonyms=b;  
 CC Note=Gene prediction confirmed by EST data;  
 CC Name=3; Synonyms=c;







Db 484 GESASPAN 491

RESULT 12

THN2\_VISAL STANDARD; PRT; 46 AA.

AC P32880; P01536;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Viscotoxin A2.

GN TH12.3.

OS Viscum album (European mistletoe).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Santalales; Santalaceae; Viscum.

OX NCBI\_TaxID=3972;

[1]

RP SEQUENCE.

RX MEDLINE=72211843; PubMed=5035954;

RA Olson T., Samuelsson G.;

RT "The amino acid sequence of viscotoxin A2 from the European mistletoe (Viscum album L., Loranthaceae).";

RL Acta Chem. Scand. 26:585-595(1972).

[2]

RP DISULFIDE BONDS.

RX MEDLINE=75013480; PubMed=4607177;

RA Olson T., Samuelsson G.;

RT "The disulphide bonds of viscotoxin A2 from the European mistletoe (Viscum album L. Loranthaceae).";

RL Acta Pharm. Suec. 11:381-386(1974).

CC -!- FUNCTION: Thionins are small plant proteins which are toxic to animal cells. They seem to exert their toxic effect at the level of the cell membrane. The precise function, in plants, of these proteins is not known.

CC -!- SIMILARITY: Belongs to the plant thionin family.

DR PIR; A90005; VTVA2.

DR HSSP; P01544; 2PLH.

DR InterPro; IPR001010; Thionin.

DR Pfam; PF00321; plant thionins; 1.

DR PROSITE; PS00271; THIONIN; 1.

KW Plant defense; Thionin; Plant toxin.

FT DISULFID 3 40 PROBABLE.

FT DISULFID 4 32

FT DISULFID 16 26

SQ SEQUENCE 46 AA; 4834 MW; 61149207EE3FF0E5 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 46;

Best Local Similarity 100.0%; Pred. No. 9.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 NTCRFGG 18

DB 14 NTCRFGG 20

RESULT 13

THN\_PHOTO STANDARD; PRT; 46 AA.

AC P01539;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phoratoxin.

OS Phoradendron tomentosum (California mistletoe).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Santalales; Santalaceae; Phoradendron.

OX NCBI\_TaxID=3969;

[1]

RP SEQUENCE.

RX STRAIN-Subsp. macrophyllum;

RX MEDLINE=75013484; PubMed=4415051;

Db 484 GESASPAN 491

RESULT 14

SUGB\_CITFR STANDARD; PRT; 105 AA.

AC O69279;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SUGB protein homolog.

GN SUGB.

OS Citrobacter freundii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI\_TaxID=546;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=OS60;

RX MEDLINE=98344100; PubMed=9677290;

RA Bishop R.E., Leski B.K., Hodges R.S., Kay C.M., Weiner J.H.;

RT "The entericidin locus of Escherichia coli and its implications for programmed bacterial cell death.";

RL J. Mol. Biol. 280:563-596(1998).

CC -!- FUNCTION: COULD ACT AS A CHAPERONE (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: Belongs to the small multidrug resistance (SMR) protein family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

RA Mellstrand S.T., Samuelsson G.;

RT "Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. macrophyllum (Loranthaceae). The amino acid sequence.";

RL Acta Pharm. Suec. 11:347-360(1974).

RN [2]

RP SEQUENCE.

RA Samuelsson G.;

RT "Mistletoe toxins.";

RL Syst. Zool. 22:566-569(1973).

RN [3]

RP DISULFIDE BONDS.

RX MEDLINE=75013480; PubMed=4606908;

RA Mellstrand S.T., Samuelsson G.;

RT "Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. macrophyllum (Loranthaceae). The disulphide bonds.";

RL Acta Pharm. Suec. 11:367-374(1974).

CC -!- FUNCTION: Thionins are small plant proteins which are toxic to animal cells. They seem to exert their toxic effect at the level of the cell membrane. The precise function, in plants, of these proteins is not known.

CC -!- SIMILARITY: Belongs to the plant thionin family.

DR PIR; A01802; VFPL1T.

DR HSSP; P01544; 2PLH.

DR InterPro; IPR001010; Thionin.

DR Pfam; PF00321; plant thionins; 1.

DR PROSITE; PS00271; THIONIN; 1.

KW Plant defense; Thionin; Plant toxin.

FT DISULFID 3 40 PROBABLE.

FT DISULFID 4 32

FT DISULFID 16 26

FT MOD\_RES 46 46

SQ SEQUENCE 46 AA; 4881 MW; A02A08BA4E27D1A9 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 46;

Best Local Similarity 100.0%; Pred. No. 9.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 NTCRFGG 18

DB 14 NTCRFGG 20

RESULT 14

SUGB\_CITFR STANDARD; PRT; 105 AA.

AC O69279;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SUGB protein homolog.

GN SUGB.

OS Citrobacter freundii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI\_TaxID=546;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=OS60;

RX MEDLINE=98344100; PubMed=9677290;

RA Bishop R.E., Leski B.K., Hodges R.S., Kay C.M., Weiner J.H.;

RT "The entericidin locus of Escherichia coli and its implications for programmed bacterial cell death.";

RL J. Mol. Biol. 280:563-596(1998).

CC -!- FUNCTION: COULD ACT AS A CHAPERONE (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: Belongs to the small multidrug resistance (SMR) protein family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U21727; AAC46457.1; -
DR InterPro: IPR000390; Smr. Drug Res; 1.
DR Pfam: PF00893; Multi_Drug_Res; 1.
KW Chaperone; Transmembrane; Transport.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
SQ SEQUENCE 105 AA; 10897 MW; 62DA129ADA86E765 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 GESASPA 734
Db 79 GESASPA 85

RESULT 15
ACPS BRAJA
ID ACPS BRAJA STANDARD; PRT; 139 AA.
AC O89159;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE (4'-phosphopantetheinyl transferase acps).
CN ACPS OR BLL5063.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110SPC4;
RX MEDLINE=95086246; PubMed=9870699;
RA Baird A., Mueller P.;
RT "A second gene for type I signal peptidase in Bradyrhizobium
RT japonicum, sipF, is located near genes involved in RNA processing and
RT cell division.";
RL Mol. Gen. Genet. 260:346-356(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
CC -!- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
CC A to a Ser of acyl-carrier protein (By similarity).
CC -!- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
CC 3',5'-bisphosphate + holo-[acyl-carrier protein].
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the P-pant transferase superfamily. Acps
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF065159; AAD02937.1; -
DR EMBL: AP005953; BAC50328.1; -
DR HAMAP: MF_00101; -; 1.
DR InterPro: IPR008278; 4-PPT_transf.
DR InterPro: IPR002582; ACPS_transf.
DR Pfam: PF01648; ACPS; 1.
DR ProDom: PD004282; ACPS; 1.
DR TIGRFAMs: TIGR00516; acps; 1.
DR TIGRFAMs: TIGR00556; pantethn_trn; 1.
KW Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
KW Complete proteome.
FT METAL 8 8 MAGNESIUM (BY SIMILARITY).
FT METAL 61 61 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 139 AA; 15193 MW; 2FD4A315845FC788 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 522 ARIDVSI 528
Db 112 ARIDVSI 118

Search completed: August 25, 2004, 18:12:25
Job time : 35 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:26:03 ; Search time 27 Seconds  
(without alignments)  
1496.535 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 4078  
Sequence: 1 MTGIAAASFFNTCRFGCG.....KALETDSVSGVSKSQKL 776

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3413	83.7	739	1 JUAZ HUMAN	Q15022 homo sapien
2	3332	81.7	741	1 JUAZ MOUSE	Q80u70 mus musculu
3	1127.5	27.6	900	1 SUI12 DROME	Q9njg9 drosophila
4	661	16.2	243	1 JAZ1 HUMAN	Q85vz6 homo sapien
5	661	16.2	243	1 JAZ1 MOUSE	Q80zc5 mus musculu
6	224	5.5	631	1 EMF2 ARATH	Q816y4 arabidopsis
7	217	5.3	440	1 VRN2 ARATH	Q8w5b1 arabidopsis
8	211	5.2	692	1 FIS2 ARATH	Q9znt9 arabidopsis
9	176.5	4.3	1142	1 GIN4 YEAST	Q12263 saccharomyc
10	161	3.9	1453	1 Y373 BOVIN	Q9tu23 bos taurus
11	155.5	3.8	1755	1 YJ27 YEAST	P47098 saccharomyc
12	152	3.7	1957	1 SPOF SCHPO	Q10411 schizosacch
13	150	3.7	912	1 UBP3 YEAST	Q01477 saccharomyc
14	147.5	3.6	1790	1 USO1 YEAST	P25386 saccharomyc
15	147	3.6	2230	1 GOA4 HUMAN	Q13439 homo sapien
16	146.5	3.6	1235	1 TRK1 YEAST	P12685 saccharomyc
17	145	3.6	1744	1 TANA XENLA	Q01550 xenopus lae
18	145	3.6	1953	1 BNI1 YEAST	P41832 saccharomyc
19	143.5	3.5	698	1 YB06 YEAST	P38283 saccharomyc
20	141.5	3.5	1328	1 YMD9 YEAST	Q03434 saccharomyc
21	141	3.5	1875	1 MUP1 YEAST	Q02455 saccharomyc
22	139.5	3.4	1085	1 YAP4 SCHPO	Q09863 schizosacch
23	139.5	3.4	1833	1 ZEP2 HUMAN	P31629 homo sapien
24	137	3.4	1969	1 Z292 HUMAN	O60281 homo sapien
25	136	3.3	2867	1 RBP2 PLAVB	Q00799 plasmodium
26	135.5	3.3	1544	1 ARHC HUMAN	Q9nzn5 homo sapien
27	135	3.3	3210	1 CENF HUMAN	P49454 homo sapien
28	135	3.3	5430	1 MACF HUMAN	Q9upn3 homo sapien
29	135	3.3	5938	1 MAC4 HUMAN	Q96pk2 homo sapien
30	134.5	3.3	1148	1 RF01 HUMAN	P35251 homo sapien
31	134.5	3.3	1972	1 BA2B HUMAN	Q9uif8 homo sapien
32	133.5	3.3	1328	1 YME4 YEAST	Q04711 saccharomyc
33	132	3.2	2104	1 MYS3 SCHPO	O14157 schizosacch

ALIGNMENTS

RESULT 1

ID	JUAZ HUMAN	STANDARD;	PRT;	739 AA.
AC	Q15022; Q96BD9;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Polycomb protein SUZ12 (Suppressor of zeste12 protein homolog)			
DE	(joined to JAZF1 protein) (Chromatin precipitated E2f target 9 protein).			
GN	SUZ12 OR JUAZ1 OR CHET9 OR KIAA0160.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
RX	MEDLINE=96127530; PubMed=8590280;			
RA	Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;			
RT	"Prediction of the coding sequences of unidentified human genes. IV.			
RT	The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by			
RL	analysis of cDNA clones from human cell line KG-1."			
RL	DNA Res. 2:167-174(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Uterus;			
RX	MEDLINE=2238257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,			
RA	Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Kzywinski M.I., Skalska U., Smailus D.E.,			
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[3]			
RP	SEQUENCE OF 102-113; 343-360; 474-483 AND 509-516, IDENTIFICATION IN A			
RP	PRC2 COMPLEX WITH EED; EZH2; RBBP4 AND RBBP7, AND METHYLTRANSFERASE			
RP	ACTIVITY OF THE COMPLEX.			
RX	MEDLINE=22300550; PubMed=12351676;			
RA	Cao R., Wang L., Wang H., Xia L., Erdjument-Bromage H., Tempst P.,			
RA	Jones R.S., Zhang Y.;			
RT	"Role of histone H3 lysine 27 methylation in Polycomb-group			

Q04670 saccharomyc  
P54697 dictyosteli  
Q96t58 homo sapien  
Q12955 homo sapien  
P04932 plasmodium  
P04933 plasmodium  
Q03661 saccharomyc  
P15619 drosophila  
P47100 saccharomyc  
Q92056 mus musculu  
P46100 homo sapien  
Q7yqm4 pan troglod

silencing.";  
 Science 296:1039-1043 (2002).  
 [4]  
 RX DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH JAZF1.  
 RX MEDLINE=21265470; PubMed=11371647;  
 RA Koontz J.I., Soreng A.L., Nucci M., Kuo P.C., Pauwels P.,  
 RA van Den Berghe H., Cin P.D., Fletcher J.A., Sklar J.;  
 RT "Frequent fusion of the JAZF1 and JAZ2 genes in endometrial stromal  
 tumors.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353 (2001).  
 [5]  
 RX MEDLINE=21444978; PubMed=11564866;  
 RA Weinmann A.S., Bartley S.M., Zhang T., Zhang M.Q., Farnham P.J.;  
 RT "Use of chromatin immunoprecipitation to clone novel E2F target  
 promoters.";  
 RT Mol. Cell. Biol. 21:6820-6832 (2001).  
 [6]  
 RP IDENTIFICATION IN A PRC2 COMPLEX WITH EED; EZH2; RBBP4 AND RBBP7, AND  
 RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
 RX MEDLINE=22323099; PubMed=12435631;  
 RA Kuzmichev A., Nishio K., Erdjument-Bromage H., Tempst P.,  
 RA Reinberg D.;  
 RT "Histone methyltransferase activity associated with a human  
 multiprotein complex containing the Enhancer of Zeste protein.";  
 RT Genes Dev. 16:2893-2905 (2002).  
 RL  
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by  
 CC forming multiprotein complexes, which are required to maintain the  
 CC transcriptionally repressive state of homeotic genes throughout  
 CC development. PcG proteins are not required to initiate repression,  
 CC but to maintain it during later stages of development. They  
 CC probably act via the methylation of histones, rendering chromatin  
 CC heritably changed in its expressibility. Component of the PRC2  
 CC complex, which methylates Lys-9 and Lys-27 residues of histone H3.  
 CC  
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,  
 CC RBBP4 and RBBP7. This complex is distinct from the  
 CC PRC1 complex, which contains many other PcG proteins like BMI1,  
 CC CBX4, CBX6, PHC2, PHC3, SCMH1, RING1 and RNF2.  
 CC  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC  
 CC -!- INDUCTION: Induced by E2F1 transcription factor.  
 CC  
 CC -!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal  
 CC rearrangement involving JAZF1 and SUZ12. A t(7;12)(p15;q21)  
 CC chromosomal translocation involving JAZF1 and SUZ12 generates the  
 CC JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1  
 CC and the C-terminus part of SUZ12. The translocation is frequently  
 CC found in all cases of endometrial stromal tumors, except in  
 CC endometrial stromal sarcomas, where it is rarer.  
 CC  
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(Z)12) family.  
 CC  
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcements> or  
 CC send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC  
 DR EMBL; D63881; BAA09931.1; ALT INIT.  
 DR EMBL; BC015704; AAH15704.1; -;  
 DR MIM; 606245; -;  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1;  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; FALSE\_NEG.  
 KW Transcription regulation; Repressor; Chromatin regulator;  
 KW Nuclear protein; Metal-binding; Zinc; Zinc-finger;  
 KW Chromosomal translocation; Proto-oncogene.  
 FT ZN FING 448 471 C2H2-TYPE.  
 FT DOMAIN 563 639 VEFs-BOX.  
 FT DOMAIN 7 50 GLY-RICH.  
 FT DOMAIN 51 59 POLY-SER.  
 FT DOMAIN 60 67 POLY-ALA.  
 FT SITE 93 94 BREAKPOINT FOR TRANSLOCATION TO FORM

FT CONFLICT 216 216 JAZF1-SUZ12 ONCOGENE.  
 SQ SEQUENCE 739 AA; 83053 MW; A8930BBD031DF656 CRC64;  
 Query Match 83.7%; Score 3413; DB 1; Length 739;  
 Best Local Similarity 99.8%; Pred. No. 5.6e-199;  
 Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 129 EPTQYRFLRNLNLAIPFLHRTLYMSHRSRTNKKTKFVDDMLSKVEMKGEQESH 188  
 Db :|||||  
 QY 92 KPTQYRFLRNLNLAIPFLHRTLYMSHRSRTNKKTKFVDDMLSKVEMKGEQESH 151  
 QY 189 SLSAHLQLTFTGFFHKNKPSNSNEQNSVTLVLLVKKVCHKKKDVSCPIROVPTCKK 248  
 Db :|||||  
 QY 152 SLSAHLQLTFTGFFHKNKPSNSNEQNSVTLVLLVKKVCHKKKDVSCPIROVPTCKK 211  
 QY 249 QVPLIPDLNQTGKGNFPSPSLAVSSNEFEFSPNSHVKSYSLFRVTRPGRRENGMNGETN 308  
 Db :|||||  
 QY 212 QVPLIPDLNQTGKGNFPSPSLAVSSNEFEFSPNSHVKSYSLFRVTRPGRRENGMNGETN 271  
 QY 309 ENIDVNEELPARKKRNREDGKTFVAQMTVPDKNRLQLLDGEYEVAMQEMEECPISKKR 368  
 Db :|||||  
 QY 272 ENIDVNEELPARKKRNREDGKTFVAQMTVPDKNRLQLLDGEYEVAMQEMEECPISKKR 331  
 QY 369 ATWETILDGKRLPPPTETFSQGTPLQFTLRWTGTTNDKSTAPIAKPLATRNSESILHQENKP 428  
 Db :|||||  
 QY 332 ATWETILDGKRLPPPTETFSQGTPLQFTLRWTGTTNDKSTAPIAKPLATRNSESILHQENKP 391  
 QY 429 GSVKPTQTIAVKSITTDLQTRKXDTNENKQLRIFQFLYNNNTTQCTEARDLHCP 488  
 Db :|||||  
 QY 392 GSVKPTQTIAVKSITTDLQTRKXDTNENKQLRIFQFLYNNNTTQCTEARDLHCP 451  
 QY 489 WCTLNCRKLYSLKHLKLSHRSFIENYVYHPKARIDVSIENECYDGSVAGNPQDIHRQPG 548  
 Db :|||||  
 QY 452 WCTLNCRKLYSLKHLKLSHRSFIENYVYHPKARIDVSIENECYDGSVAGNPQDIHRQPG 511  
 QY 549 FAFSRNGPVKPTITHLVCRPKRTKASMSPELSEDEGEVQOORTYSSGHNLRFHSDTC 608  
 Db :|||||  
 QY 512 FAFSRNGPVKPTITHLVCRPKRTKASMSPELSEDEGEVQOORTYSSGHNLRFHSDTC 571  
 QY 609 LPLRQCEMEVDSDEKDPFWLREKTIITQIEFSDVNEGEKEVKMLNHLVHKHGFADNQ 668  
 Db :|||||  
 QY 572 LPLRQCEMEVDSDEKDPFWLREKTIITQIEFSDVNEGEKEVKMLNHLVHKHGFADNQ 631  
 QY 669 MNHACMLFVENVGQKIKNKLCRNFMHLVSHMDFNLISIMSIDKAVTKLREMOKLEBK 728  
 Db :|||||  
 QY 632 MNHACMLFVENVGQKIKNKLCRNFMHLVSHMDFNLISIMSIDKAVTKLREMOKLEBK 691  
 QY 729 ESASPAHEITEFQNGTANGSEINSKKALETDSVGSVSKQSKQL 776  
 Db :|||||  
 QY 692 ESASPAHEITEFQNGTANGSEINSKKALETDSVGSVSKQSKQL 739  
 RESULT 2  
 ID JAZ2\_MOUSE STANDARD; PRT; 741 AA.  
 AC Q80U70; Q80Y10; Q99L07;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Polycomb protein Suz12 (Suppressor of zeste 12 protein homolog).  
 GN SUZ12 OR KIAA0160 OR D11ERTD530E.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22579291; PubMed=13693553;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,  
 RA Nakajima D., Nagase T., Ohara O., Koga H.;  
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:



RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA April J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.S., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriter S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kratt C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RN REVISIONS, AND ALTERNATIVE SPLICING.  
 RP MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kinkler J.S., Millburn G.H., Prochman S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RN SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=Berkeley; TISSUE=Embryo;  
 RA MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Kronmiller B., Paclebb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A *Drosophila* full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RN [5]  
 RN IDENTIFICATION IN A ESC/E(2) COMPLEX WITH B(2); CAF1 AND ESC,  
 RP AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
 RX MEDLINE=22296673; PubMed=12408863;  
 RA Czermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V.,  
 RT "Drosophila enhancer of Zeste/ESC complexes have a histone H3  
 RT methyltransferase activity that marks chromosomal Polycomb sites.";  
 RL Cell 111:185-196(2002).  
 RN [6]  
 RN IDENTIFICATION IN A ESC/E(2) COMPLEX WITH B(2); CAF1 AND ESC, AND  
 RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
 RX MEDLINE=22296674; PubMed=12408864;  
 RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,

RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;  
 RT "Histone methyltransferase activity of a *Drosophila* Polycomb group  
 RT repressor complex.";  
 RL Cell 111:197-208(2002).  
 CC -i- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by  
 CC forming multiprotein complexes, which are required to maintain the  
 CC transcriptionally repressive state of homeotic genes throughout  
 CC development. PcG proteins are not required to initiate repression,  
 CC but to maintain it during later stages of development. They  
 CC probably act via the methylation of histones, rendering chromatin  
 CC heritably changed in its repressibility. Component of the Esc/E(z)  
 CC complex, which methylates Lys-9 and Lys-27 residues of histone H3.  
 CC Despite the presence of a zinc-finger, it does not bind directly  
 CC to DNA, the Esc/E(z) complex being probably recruited to DNA by  
 CC Pho. The Esc/E(z) complex is necessary but not sufficient to  
 CC recruit a functional PcG repressive complex that represses target  
 CC genes, suggesting that the recruitment of the distinct PRC1  
 CC complex is also required to allow a subsequent repression.  
 CC -i- SUBUNIT: Component of the Esc/E(z) complex, composed of Esc, E(z),  
 CC Su(z)12, Rpd3, Caf1 and probably Pho. This complex is distinct  
 CC from the PRC1 complex, which contains many other PcG proteins like  
 CC Pc, Ph, Psc, Su(z)2. The two complexes however cooperate and  
 CC interact together during the first 3 hours of development to  
 CC establish PcG silencing.  
 CC -i- SUBCELLULAR LOCATION: Nuclear.  
 CC -i- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=B;  
 CC IsoId=Q9NUG9-1; Sequence=Displayed;  
 CC Name=2; Synonyms=A;  
 CC IsoId=Q9NUG9-2; Sequence=VSP\_007033; VSP\_007034;  
 CC Note=No experimental confirmation available;  
 CC -i- SIMILARITY: Belongs to the VEPs (VRN2-EMF2-FIS2-SU(Z)12) family.  
 CC -i- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF149047; AAF73149.1; -;  
 CC EMBL; AE003515; AAF49094.2; -;  
 CC EMBL; AE003515; RAN11641.1; -;  
 CC EMBL; AY069809; RAL39954.1; -;  
 CC Flybase; FBgn0020887; Su(z)12.  
 CC GO; GO:0000790; C:nuclear chromatin; NAS.  
 CC GO; GO:0003677; F:DNA binding; NAS.  
 CC GO; GO:0016458; P:gene silencing; NAS.  
 CC GO; GO:0016481; P:negative regulation of transcription; NAS.  
 CC InterPro; IPR007087; Znf\_C2H2;  
 CC SMART; SM00355; Znf\_C2H2; 1;  
 CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1;  
 CC PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
 CC Transcription regulation; Repressor; Developmental protein;  
 CC Nuclear protein; Metal-binding; Zinc; Zinc-finger;  
 CC Alternative splicing.  
 CC ZN\_RING 411 434 C2H2-TYPE.  
 CC DOMAIN 527 603 VEPs-BOX.  
 CC DOMAIN 355 366 ASN-RICH.  
 CC DOMAIN 699 879 SER-RICH.  
 CC VARSFLIC 806 855  
 CC VARSFLIC 856 900 Missing (in isoform 2).  
 CC MUTAGEN 274 274 G->D: IN SU(Z)12-2; INDUCES LARVAL  
 CC SEQUENCE 900 AA; 100104 MW; 53BA0D83C49BC92F CRC64;



Query Match		27.6%; Score 1127.5; DB 1; Length 900;
Best Local Similarity		32.9%; Pred. No. 9.3e-61;
Matches	273; Conservative	137; Mismatches 240; Indels 179; Gaps 26;
QY	65	MTDAARREOESLKKIQPKLSLTGLASSVSRGNV-----STPPRHSSGSLTPPVTPPTSS 120
DB	1	MAPAKKREKDS-----NPDGSAANGIIGLTHGAPDASNAGSTVPTPAEQVKLN 49
QY	121	SFSSSTP-----TEPTQIYRLRNLAPIFLHRTLYMSHRNSRNIRKKTPEVDDM 174
DB	50	GHOQOEQLFQAFKPTQIYRLRNLHETNPFLARTLSYKERNRNKRKRIISQVNSM 109
QY	175	LSKVKMKGEQESHLSA-HLQITFTGFHKDKSPSENQ-----NSVTE 222
DB	110	LESI-----TQSEAVSONYLHVYDLSLHEKLPARLDNESGDLLOEQLLCEAGSVSVE 164
QY	223	VLLVVKCHKRKDVCPTRQVTKQKQVPLIPDLNQTQPGNPPSLAVSNFEP--SNSHM 281
DB	165	TLYKITSRKDKDSTLDFOELLSKCSQIVYNP---KDRVGEHATISPIQTMRPMEGHT 221
QY	282	VKSYSLLFRVTRPGRREFNGMNGTENNIDVNEELPARKKENREDGKTFVAQMTVFDK 341
DB	222	L--YKLLFRK-----VLSPTCN--DNAETPP--NKSRPN--EMFGSEILILEYK 266
QY	342	NRRLQDGEYEVAMQEMECPI---SKKRAWTETILDKRLP---PPTFSQGETLOFT 395
DB	267	SSGF--ITEGEYEAQLPLNSTSIKFSFKCTWETMPD--SYIPLSLTYDVVQSPMLKXF 324
QY	396	LWGTETNDKSTAPIAKPLATR-----NSESLOEN-----KGSVKPTQTI 437
DB	325	LTL---SNQLEPMSIAPELQYVQHLDAVEMNMYNNNNNNNCSGLKNGSGGNTV 361
QY	438	AVKESLTTDLQTRKEKDTNENRQKRLIPYQLYNNTRQOEARDLHCPCWCTLNCRKL 497
DB	382	C-----KTPP-----EHIQIYVFNWYNNTRQTEYTOELNCPWGLDCLRL 423
QY	498	YSLKHLKLCRGRPIFNYPYHKGARIDVSNECVDSYAGNPQDIHQPGAFPSRN--GP 556
DB	424	YALLKHLKLCRGRPIFNYPYHKGARIDVINDAYDGSYAGNPQDIHQPGAFPSRN--GP 483
QY	557	VKRTPTHTLVCRPKTKASMEFLESDEGEVEQRTYSSGHNRLYFHSDFCLPLRPOEM 616
DB	484	VRTSVTLVCRPKTKASMEFLESDEGEVEQRTYSSGHNRLYFHSDFCLPLRPOEM 543
QY	617	EYDSEKDPWLREKTIQIEFSDVNEGEKVMKLNHVMKGFADQNMHACMLF 676
DB	544	DIDSGESDPLRQKTIQIMIDEFSDVNEGEKVMKLNHVMKGFADQNMHACMLF 603
QY	677	VENYQKLIKKNLCLRNFLHLYSMDFNLISIMSDKAVTKLREMOQKLEKES----- 730
DB	604	LDKAGTEIVRKNLYRNFLHLMCSLFDYGLTAAETVYKTVQKLGLLSKYAAGQELMQOR 663
QY	731	-----ASPAEEITE----- 740
DB	664	BEQLXWLDVGMHKKQEDPKTKSPQKAPPADQASTSSATSGSGSSSSNQPKKMPA 723
QY	741	-----FONGTANGSEINSKEKALETSVSGVSKQKQKL 776
DB	724	HLKRGSAASSPGVQSGKTENGTSNGNS-----SSNSKNVAKSADQPL 767

RESULT 4

JAZ1_HUMAN				
ID	JAZ1_HUMAN	STANDARD;	PRT;	243 AA.
AC	Q86VZ6; Q8N3L7;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Juxtaposed with another zinc finger protein 1.			
GN	JAZF1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			

OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A. (ISOFORM 1).
RC	TISSUE=Brain, and Testis;
RX	MEDLINE=22388257; PubMed=12477932;
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA	Altschul S.F., Zeeberg B., Buetow K.H., Shenmen C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raba S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA	Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Whiting J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Shevchenko Y., Bouffard G.G.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA	Butterfield Y.S.N., Kzywinski M.I., Skalska U., Smailus D.E.,
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]	
RP	PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
RC	TISSUE=Amalgam;
RA	Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
RA	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[3]	
RP	DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH SUZ12.
RX	MEDLINE=21265470; PubMed=11371647;
RA	Koontz J.I., Sorong A.L., Nucci M., Kuo F.C., Pauwels P.,
RA	van Den Berghe H., Cin P.D., Fletcher J.A., Sklar J.;
RT	"Frequent fusion of the JAZF1 and JUAZ1 genes in endometrial stromal
RT	tumors."
RL	Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353(2001).
CC	-!- FUNCTION: Potential transcription factor.
CC	-!- SUBCELLULAR LOCATION: Nuclear protein (Potential).
CC	-!- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Names=1;
CC	Isoid=Q86VZ6-1; Sequence=Displayed;
CC	Names=2;
CC	Isoid=Q86VZ6-2; Sequence=VSP_007755;
CC	Note=Sequence incomplete. No experimental confirmation
CC	available;
CC	-!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal
CC	rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21)
CC	chromosomal translocation involving JAZF1 and SUZ12 generates the
CC	JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1
CC	and the C-terminus part of SUZ12. The translocation is frequently
CC	found in all cases of endometrial stromal tumors, except in
CC	endometrial stromal sarcomas, where it is rarer.
CC	-!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	at the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
CC	EMBL; BC042441; AAH42441.1; --
CC	EMBL; AL834234; CAD38912.1; --
DR	MIM; 606246; --
DR	InterPro; IPR007087; Znf_C2H2.
DR	Pfam; PF00096; zf-C2H2; 3.
DR	SMART; SM00355; Znf_C2H2; 3.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR	PROSITE; PS01517; ZINC_FINGER_C2H2_2; FALSE_NEG.

KW Transcription regulation; Nuclear protein; Metal-binding; Zinc;  
 KW Zinc-finger; Chromosomal translocation; Proto-oncogene;  
 KW Alternative splicing;  
 FT ZN\_FING 12 37 C2H2-TYPE 1.  
 FT ZN\_FING 173 198 C2H2-TYPE 2.  
 FT ZN\_FING 208 230 C2H2-TYPE 3 (DEGENERATE).  
 FT SITE 129 130 BREAKPOINT FOR TRANSLLOCATION TO FORM  
 FT JAZF1-SUZ12 ONCOGENE.  
 FT VARSPLIC 1 38 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNH1 ->  
 FT YFVLGSLYSDAKKRNCVFT (in isoform 2).  
 FT CONFLICT 133 133 /FTIQ=VSP 007755.  
 FT D -> G (IN REF. 1).  
 FT SEQUENCE 243 AA; 27079 MW; 878A6EE3D82C0588 CRC64;  
 Query Match 16.2%; Score 661; DB 1; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-33;  
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60  
 DB 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60  
 QY 61 INRFMTDAARBOESLKKIOPKSLTSSVSRSRGVSTPPRHSSGSLTPVTPPTPSS 120  
 DB 61 INRFMTDAARBOESLKKIOPKSLTSSVSRSRGVSTPPRHSSGSLTPVTPPTPSS 120  
 QY 121 SPFSSTPT 128  
 DB 121 SPFSSTPT 128  
 RESULT 5  
 ID JAZ1 MOUSE STANDARD; PRT; 243 AA.  
 AC Q80Z05; Q8BLQ8; Q8BWN4;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Juxtaposed with another zinc finger protein 1.  
 GN JAZF1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=C57BL/6J; TISSUE=Brain cortex, and Pancreas;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bragdt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragoti T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,  
 RA Grimond S., Gusninch S., Hirokawa N., Jackson I.J., Jarvis B.D.,  
 RA Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertege G.,  
 RA Petkovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Vetrardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imorani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 RP [2]  
 RC SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Shat N.K.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hsieh F.,  
 RA Spaplen M., Soares M.B., Bonaldi M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska J., Smalios D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -!- FUNCTION: Potential transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear protein (Potential).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoID=Q80ZQ5-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoID=Q80ZQ5-2; Sequence=VSP 007756;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AK043767; BAC31649.1; -;  
 CC EMBL; AK050474; BAC34275.1; -;  
 CC EMBL; BC048577; AAH48577.1; -;  
 CC InterPro; IPR007087; Znf\_C2H2.  
 CC Pfam; PF00096; zf-C2H2; 3.  
 CC SMART; SM00355; Znf\_C2H2; 3.  
 CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2..  
 CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
 CC Transcription regulation; Nuclear protein; Metal-binding; Zinc;  
 CC Zinc-finger; Alternative splicing.  
 FT ZN\_FING 12 37 C2H2-TYPE 1.  
 FT ZN\_FING 173 198 C2H2-TYPE 2.  
 FT ZN\_FING 208 230 C2H2-TYPE 3 (DEGENERATE).  
 FT VARSPLIC 64 64 Missing (in isoform 2).  
 FT /FTIQ=VSP 007756.  
 FT CONFLICT 211 211 R -> L (IN REF. 1; BAC34275).  
 FT SEQUENCE 243 AA; 27097 MW; 878A6EE3D82C0348 CRC64;  
 Query Match 16.2%; Score 661; DB 1; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-33;  
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60  
 DB 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60







transcriptionally repressive state of homeotic genes throughout development. PCG proteins are not required to initiate repression, but to maintain it during later stages of development. They probably act via the methylation of histones, rendering chromatin heritably changed in its expressibility. Required to prevent the proliferation of the central cell by repressing unknown target genes before fertilization.

CC -!- SUBUNIT: Probably indirectly associated with FIE and/or MEA. In plants, PCG complexes are probably composed of a member of the EZ family (CLF or MEA), FIE, and a member of the VEFs family (FIS2, VRN2 or EMF2) (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: Weakly expressed. Expressed in late siliques.

CC -!- DEVELOPMENTAL STAGE: Expressed maternally and zygotically. Expressed in the central cell before fertilization, and in the endosperm after fertilization, then decreases before the time of endosperm cellularization but continues in the chalazal cyst.

CC -!- POLYMORPHISM: In cv. Columbia, the sequence differs from that shown due to a deletion in the genomic sequence that covers the Leu-443 to Ile-502 region.

CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(Z)12) family.

CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF096095; AAD09104.1; -.  
 CC EMBL; AF096096; AAD09105.1; -.  
 CC EMBL; AF096098; AAD15448.2; ALT\_SEQ.  
 CC InterPro; IPR007087; Znf\_C2H2.  
 CC SMART; SM00355; Znf\_C2H2; 1.  
 CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 CC PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
 CC Transcription regulation; Repressor; Nuclear protein;  
 KW Developmental protein; Repeat; Zinc-finger; Metal-binding;  
 KW Polymorphism.  
 FT ZN-FING 11 34  
 FT DOMAIN 122 502  
 FT REPEAT 122 143  
 FT REPEAT 161 183  
 FT REPEAT 184 206  
 FT REPEAT 207 228  
 FT REPEAT 229 250  
 FT REPEAT 259 280  
 FT REPEAT 299 320  
 FT REPEAT 338 360  
 FT REPEAT 361 381  
 FT REPEAT 399 421  
 FT REPEAT 422 442  
 FT REPEAT 459 480  
 FT REPEAT 481 502  
 FT DOMAIN 144 519  
 FT REPEAT 144 160  
 FT REPEAT 251 267  
 FT REPEAT 282 298  
 FT REPEAT 321 337  
 FT REPEAT 382 398  
 FT REPEAT 443 458  
 FT REPEAT 503 519  
 FT DOMAIN 543 621  
 FT VARIANT 269 298  
 FT VARIANT 442 501  
 FT VARIANT 554 554  
 FT MUTAGEN 637 692  
 FT SEQUENCE 692 AA; 77666 MW; 2A3F61B961676497 CRC64;  
 Query Match 5.2%; Score 211; DB 1; Length 692;

Best Local Similarity 17.8%; Pred. No. 1.6e-05;  
 Matches 149; Conservative 122; Mismatches 286; Indels 278; Gaps 30;

QY 19 CGLFPTLADLIEHEDNH-----IDTPRV-----LKKQELQQPTY 55  
 DB 20 CGGH-----BGLQLHKKSHDAKFFBYRAEKDHGPEVDVSVKSDTIKFGVLKDDVGNPQL 75

QY 56 VALSYINRFTDAAREQESLKKIKQ-----PKLSLTSSSVSRGNVSTPR-H 103  
 DB 76 SPULFCCKNRNRRORDDNNVKLVNLMELDDLDPGRGTENDSTHVDNVSFPRAH 135

QY 104 SSGSLTPPTPTPTSSSRSSSTPTPTQIVAFLETRNLIAPIFLERTLTMYSHNRSTN 163  
 DB 136 SSEKISDILT--TTQLAIAESSEPKV-----HVNDGNVSGPPRAHS---AEKNSTH 184

QY 164 IRRKFKVDDMLSKVEKMGESLSNAHLQLTITGFFHKNDKSPNSN----- 214  
 DB 185 VN-----DD-----DDVSSPPRAHSL-----EKNESTHVNEDNISPPKAHS 221

QY 215 -EONSVTLEVLVVKVCHKKRDKVSCPIR-----QVPTGKQVPLIPDLNQTQFGN-- 263  
 DB 222 SKKNST-----HWNDEDVSFPPTRSSKETSDILTTQPAIVFSEPKVRRGSR 272

QY 264 -----FPLAVSS-----NEPEPSNMVKSYSLLFRVTRPG----- 295  
 DB 273 KQLYAKRYKARETOPAIAESSEPKVLHVNDENVSSPPEAHSLEKASDILTTQPAIAESS 332

QY 296 -----RRFENGINGETNENIDVNEELPARRKRNEDGEGTFAQMTVPD 340  
 DB 333 EPKPVHNDENVSSPTFAHSSKKNKSTRKVD--NVSPSPKTRSSKKTSDILTTQPTIAE 391

QY 341 KNRRLQLDGEYAVQEMEECPISKGRATWETILDGKRLPPFTFSGOFTLQFTLRWTG 400  
 DB 392 SS-----EPKVRHWNDDNVS-----TPRAHS 413

QY 401 ETNDKSTAPIAKPLATRNSESLSHQENKPGSKVKPTOTIAKESLTDTLQTRKEKDTNENR 460  
 DB 414 SKKNKST-----RKNDNIPSPPKTRSSKKTNLTTR---TOPAIAESEPKVPHVND 462

QY 461 QKLRIFYQFLYNNTRQQTAEADD--LHCPWCCTLCNCKLYSLKLKLCCHSRFIFYVYHP 519  
 DB 463 DKVSTPTP-AHSSKKNKSTHKDDNASLPPKTRSSKKTSDIL----- 503

QY 520 KGARIDVSNEDYDGSYAGNPQDIHQFGFAPSRNGPKRPIITHILVCRPKRYKASME 579  
 DB 504 -----ATTQPAKAEF-----SEPKVTRVSRK 525

QY 580 FLESDEGEVEQRTYSSGHNRLYFHSHTCLPLRPOEM--EVDSEDEKDPWLRKTTIQI 637  
 DB 526 ELHAERCEAKRLERLK--GRQFYHSQTMQPMTPFQVMSNEDSENEDDYALDLSERLRL 582

QY 638 EEFSDVNEGEKEVMKLVNLMHVMKHGFIADNQNHACMLFVE-----NYGQKII 685  
 DB 583 ERLVGVSKEEKRYMYLWNIFVRKQVRIADGHVPWACEEFAKLHKEEMKNSSFDWWRMF 642

QY 686 KKNLCNFMHLVSMHDFNLISIMSIDKAVTKLRWQOKLEKESASPANBEITE 740  
 DB 643 RIKLWNGLICAKTFHKCTTILLNSDEA-----GQFTSGSANNANNOQSM 689

RESULT 9  
 GIN4 YEAST  
 ID GIN4 YEAST STANDARD; PRT; 1142 AA.  
 AC Q1223;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Serine/threonine-protein kinase GIN4 (EC 2.7.1.-).  
 GN GIN4 OR YDR507C OR D9719.13.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetia; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;

```

[1]
SEQUENCE FROM N.A.
RP MEDLINE=99030835; PubMed=9813093;
RA Longline M.S., Fares H., Pringle J.R.;
RT "Role of the yeast Gln4p protein kinase in septin assembly and the
RL relationship between septin assembly and septin function.";
J. Cell Biol. 143:719-736(1998).
[2]
SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J., Allen E., Atauljo R., Aviles E., Berno A.,
RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Hunnicke-Smith S., Hyman R., Komp C., Laekari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA Roberts A., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Botstein D., Davis R.W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: May play a role in septin assembly.
CC !- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. NIM1
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U33140; AAA75513.1; -.
CC EMBL; U33057; AAB64949.1; -.
CC PIR; S59359; S59359.
CC HSP; O63450; 1A06.
CC GerOnline; 140999; -.
CC SGD; S0002515; GIN4.
CC GO; GO:0005935; C:bud neck; IDA.
CC GO; GO:0007117; P:bud growth; IGI.
CC GO; GO:0001135; P:septin checkpoint; IGI.
CC GO; GO:0000921; P:septin ring assembly; IGI.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR008271; Ser_thr_kin_AS.
CC InterPro; IPR002290; Ser_thr_kinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PD00109; TYRKINASE.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 19 289 PROTEIN_KINASE.
FT NP_BIND 25 33 ATP (BY SIMILARITY).
FT BINDING 48 48 ATP (BY SIMILARITY).
FT ACT_SITE 136 156 BY SIMILARITY.
SQ SEQUENCE 1142 AA; 129857 MW; EC16FF4BB49DD811 CRC64;

Query Match 4.3%; Score 176.5; DB 1; Length 1142;
Best Local Similarity 20.8%; Pred. No. 0.0037;
Matches 152; Conservative 102; Mismatches 264; Indels 213; Gaps 35;

QY 41 DRVLEKQELQEP-----TYVALSYINRFMTDAARQESLKKIQKLSLTSS-SVS 93
DB 337 DPEGI-KGKLREPGANAEKTLVALLY--RPKCDT----QKELIKQQQVKRQSSVSVS 389
QY 94 RG-NVSTPP-----RHSSGSSTPPVTPPI-----TPSSFRSSTPTEPTQIVRFLNTR 140
DB 390 PSKKVSTTPQRRNRRESLIVTSRRKKPISFNKFTASSASSNLTPGSSKRLSKNFSK 449
QY 141 NLIAPFIHLRTLYMHSNRSTNI--KSKTKPVDDMLSKVRKMGQES----- 187
DB 450 KKLSTIVNQSSPTPAS-RNKRASVINVEKNQKRASIPSTTKNKRSSRSIKRMSLIIPSMK 508
QY 188 -HLSAHLQLFTGTGFHKNDKRPSPNEQNSVTLVLLVKV-----CHKKKQDVSC 238

```

```

Db 509 RESVTKLMSTYAKLAEDDDWEYIEKTRKTSNFATLIDEIFEYKYEQIRKEBELER 568
QY 239 PIQVPTG-----KQVPL-IPDLNO----- 258
DB 569 KVREAKAREELRRRRKQBEKARKLLEKDKRKQBELKQIEIDISLEBELSKHKE 628
QY 259 -TXFGNPPSLAVSNEFEPSNHHVKSYSLLFR-----VTRGRRREFNGMINGETNENI 311
DB 629 EKLGDNIRISAPWENEENKINHLVDIDNILRRNFSLQTRPVSRLDPGIMFSSPTEEV 688
QY 312 DVNEELPARKRNREDEGKTFVQM-----TVFDKNRRLQLLDGEYE--VAMQEM-EE 361
DB 689 SPIVE--PKRTENERLTTTEKILTIIRSKPLGSGFNIDKELKSKRWEYFSIIAPQLSSE 746
QY 362 CPISKKRAATWETIL---DGKRLPPFE--TPSQOQPTLQFLRWGTGETNDKSTAPIAKPLAT 416
DB 747 RVVSDNSNDGYESILLPKDGNGVSQLKSDTATTAPVSDGLRKISE-----IRVPQFT 798
QY 417 RNESLHQNPKGVSVPKPTQTIIVKESLTLDQTRKCKDTPNENRQKRLRYQFLYNNNR 476
DB 799 RKRHFSESNNKRLSV--LSMYSTKESFTNLVDILKGNLDVNNQQSQRI----- 845
QY 477 QQTEARDLHCPWCTLNCRLKLYSLLKHLKLCHSRFTFNVYHPKGARIDVSIKEDVGSY 536
DB 846 PTPRSADD-----SEPLFE-----TYNE--EAEY 867
QY 537 AGNPQDIHR-----QPGFA--FSRNGFVKRTPITHILVCRPKRTKASMSE 579
DB 868 TGNSSNDRLYDVGDSTIKDSALKLNFAFRNGSNEAKQTDNLHLPLPPLNGDNLK 927
QY 580 FLESEDEVEQORTYS---SGHNRLYFHSHTCLPLRPOEMEYDSEDEKDPW----- 628
DB 928 -QNSQEGDQHPKIKSWIPESGSS--HTE-----KEENEKEEKKPEOHKOEDQ 976
QY 629 -LREKTIITQIE 638
DB 977 EKREKVDDME 987

RESULT 10
Y373 BOVIN
ID Y373 BOVIN STANDARD; PRT; 1453 AA.
AC Q9TU23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RA Jovov B., Ripoll P.J., Benos D.J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF176816; AAF00990.1; -.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 37 1426 COILED COIL (POTENTIAL).
SQ SEQUENCE 1453 AA; 169934 MW; 03CBA0A2A64CF4139 CRC64;

```



Query Match 3.9%; Score 161; DB 1; Length 1453;  
Best Local Similarity 19.3%; Pred. No. 0.044;  
Matches 165; Conservative 134; Mismatches 326; Indels 230; Gaps 36;

QY 50 LQOPTVALSYNRPMTDAARQESLKK-----IQKLSLTLSSVSRGNVSTPPR 102  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 103 HSSGSLTPVPPTPPSSFRSSTPTPTQYRFLRTNLIAPFLHRTLYTMHRSRT 162  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 680 ETAWDLIKQSTPPVPTNKHFLAEMEQVAEQDSSLSLV--IKLKQVSDLRQKEIT 737  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 163 NIKRTFK-----VDDMLSKVKKKB-----QESHLSAHLQ----- 195  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 738 ELKIKREFENMKLRLOENHAEVKKIKAEVEDLRLVQSQKESQKSELQTKQKANSRA 797  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 196 --LTFGTFHFKNDKPSNSEQNSVTLEVLVQ-----VCHKKRDVSCPIRQ 242  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 798 PTTMRLNVLRLKQALKEKQKALSALLEAEWTAABERLIISWTSQKELNVLVQ 857  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 243 VPT--GKQVPLIPDLQ-----TKPGNTPSLAVSNEPEPNSHVKSYSLFR 290  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 858 IVDRHTKLSQIEDLNENILKLEALTKSNRENTLTDNLNLTNQLNQKQKAYGVLR 917  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 291 VTRPGRRFNGMGETNENIDVNEELPAREKNRE-----DGEKTFVAQMTVFDKN 342  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 918 -----EKDAVDQENNELKQIKRITSLGQKPLINDKQSLIEL-----Q 957  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 343 RRLQLLDGEYVAMQEMEECPISKRA-----TWETILDGRLPPFPFSSQGT 391  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 958 KKIKLESQLEKVDKAEAMKPKKSAEIVRWEEGKQKTKIEGIRNKLKE--KEGEV 1015  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 392 LQFTLWGTGETNDSKTAPIAKPLATRNSESLHQENKPGSVKPTQTIIVKSLTLD----- 446  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1016 YILT-----KQTLTKDLFAKADKEKLTQK-----GLTVDQVMAA 1056  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 447 --LQTRKEKDPNENRQKLRIFYOFLYNNTRQTEARDLHCPWCTLNCRLKLYSLKHL 504  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1057 RVLSEKELEBKKNLDLNDISYMRSHQALPRDSVIEDHLQNKYLQ--EKLHLEKQL 1115  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 505 --KLCHSRIFNYVYHPKAGRIDVINECYDSYAGNPQDIHQPGFAPSRNGPVK--RTPI 562  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1116 SKDAYSR-----PSTSGID-----SDHYQ--REQELQRENKLSSNIELKFQLEQ 1160  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 563 THILVCRPK--RTKASMEFLESGEDGEVQOQ--RTYSSGHNRLYFHSDTCLPLRQEME 617  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1161 ANKOLPRUKQVRDLKEMCEFLKKEKAEVERKLGVRGSGS----- 1202  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 618 VDSDEKDPWLREKTIHQIEEFSDVNEGEKEVVKMLANLHVMGHGFADNQMHACMLFV 677  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1203 ---GKTIPE--LEKTIQGLMKVKEVQRENLK-----KASGILTSEKMAN----- 1245  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 678 ENYGOKI-----IKGNLCRNFMLHVS-----MHDNLSIM----- 709  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1246 EMENSKLAEKLEKLVHGLRSLHYESKAKGTIKVAENRERLKEKLEIKLKHVPEGD 1305  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 710 ---SIDKAVTKLRMOQKLEKESASPANEEITEEQNGTANGFSE---INSKEKALETD 762  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1306 ETEQGLQRELAVRLANSQLEKEELHRIEISKQNGPDSTISDPDLMEKIKOLETQ 1365  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 763 --SVGSVSKQKQKL 776  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679  
QY 1366 LRTSDMEKQHLKEEI 1380  
DB LQKXEVLYKY--QHLKAEQREIVKQHEELHLTKLEADNLSKF-----K 679

RESULT 11  
YJ27 YEAST  
ID YJ27 YEAST STANDARD; PRT; 1755 AA.  
AC P47098; P87194;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 30-MAY-2000 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transposon Tyl protein B.

TY1B OR YJR027W OR J1560.  
Saccharomyces cerevisiae (Baker's yeast).  
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1].  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9610930; PubMed=8619316;  
RA Zagulski M., Babinska B., Gromadka R., Migdalski A., Rytka J.,  
Sulicka J., Herbert C.J.;  
RT "The sequence of 24.3 kb from chromosome X reveals five complete open  
RT reading frames, all of which correspond to new genes, and a tandem  
RT insertion of a Tyl transposon.";  
RL Yeast 11:1179-1186(1995).  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z49526; CAA89553.1; --  
CC MEROPS; A11.003; YJR027W.  
CC SCD; S0003788; YJR027W.  
CC InterPro; IPR001969; Aspprotease AS.  
CC InterPro; IPR001042; Pept\_A11B\_Ty1.  
CC InterPro; IPR001584; Rve.  
CC Pfam; PF00665; rve; 1.  
CC Pfam; PF01021; TYA; 1.  
CC PROSITE; PS00141; ASP\_PROTEASE; 1.  
CC Transposable element; Hypothetical protein; Hydrolase;  
KW Aspartyl protease; ATP-binding.  
FT ACT SITE 461 461  
FT NP BIND 1631 1638 ATP (POTENTIAL).  
SQ SEQUENCE 1755 AA; 198615 MW; 04E248A7FD7596F CRC64;  
Query Match 3.8%; Score 155.5; DB 1; Length 1755;  
Best Local Similarity 22.6%; Pred. No. 0.12;  
Matches 159; Conservative 99; Mismatches 232; Indels 213; Gaps 44;

QY 38 IITDPRVL---EKQLQOPTVVALSY---INRFMTD-----AARQEQSLKKKI----- 80  
DB 881 VTTNVLVQKESRLDQFNVDALTFDEDLNRLTASQSFASNEIQESNDLINESDHF 940  
QY 81 -----QPKLSLTLSSVSRGNVSTPP-----RHSGSLTTP--VTPPTPS-- 119  
DB 941 QSDIELHPEQPR--NVLKAVSPTD-STPPSTHTEDSKRVSKTNIRAPREDVDNISNI 997  
QY 120 --SSFRSSTPTPTQYRFLRTN-----LIAPFLHRTLYTMHRSRTNKKKT 168  
DB 998 LPSKRSSTP-----QISNIESTGSGMKLVNPLAPMSQNT-----HESHAS-KSKD 1047  
QY 169 FKVDMLSKVKQKGEQSHLSAHLQTLTFTGFFKNDKPSNSENQNSVTLEVLVQ 228  
DB 1048 FRHSDSYS-----ENETNHTNVPISSTG--GTNNKTVPQISDQTE-----KRI 1089  
QY 229 CHKKK-KVVS-----CPIROVPTGKKQ-----VPLIPDLNQTKGNFPLSA 268  
DB 1090 IHRSPSIDASPPENNSHNIPKTPPTVSQNTSESIADLPL-PDLPSPSTFPD-- 1146  
QY 269 VSSNEFEPSNSHMKVSYSLLPVTRPGREFNGMGETNENIDVNEELPARRKRENG 328  
DB 1147 -PPKELPPINSHTNS-----SLGGI--GDSNAYTINS-----KKRSLEDN 1185  
QY 329 EKTTF-VAQMTVFDKRRRLQLLDGEYVAMQEMEECPISKR-----ATWETILDGRLPPF 383  
DB 1186 ETEIKVSRDITMNTKRR-----SLBPFRSKRIHLIAAVKAV---KSIKPI 1228  
QY 384 ETFSQOPTIQF--TLRWGTGETNDSKTAPIAKPLATRNSESLHQENKPGSVKPTQTIIVK 441  
DB ETEIKVSRDITMNTKRR-----SLBPFRSKRIHLIAAVKAV---KSIKPI 1228

Db 1229 RT-----TLRYDEATYNKDIKE-----KYIEAYHKE-----VNQLLKMT 1266

Qy 442 STLTDTQTRKEDTNNRQKLRIFVQFLYNNNT-----RQTEARDLHCP-----WCTL 492

Db 1267 WDTDEYDRKEID-----PKRVINSFMFKRKDRGTHKARFVAGDQHPDPTDTGMS 1320

Qy 493 NCRKLYSLKHLKCHSRFIFVYVHPKARDVINECYDGSAGNPDQIHRQGPAPS 552

Db 1321 NVHYHMTSLSLALDN---NYI---TQDIS---SAY---LYADIKEELVIRPPHGL 1369

Qy 553 RNPGRVTRTTHILVCRPKRTKASMEFLES---EDGEVEQORTYSSGHNRLYFHSDFCL 609

Db 1370 MNDKLIRLKKSHVL---KQSGANVYETIKSYLIRKOCGMEEVRGWS---VFKNSQVTI 1422

Qy 610 PLRPQEMEVDSEKDPWLREKTIQIEEFSD---VNEGEKE 649

Db 1423 CLFVDDMILFSKDLN---ANKKIITLKKQYDTRKIINGESD 1461

RESULT 12

SPOF\_SCHPO STANDARD; PRT; 1957 AA.

AC Q10411; Q9USE9;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Sporulation-specific protein 15.

GN SPO15 OR SPAC1F3.06C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.

RX MEDLINE=20107136; PubMed=10639340;

RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;

RT "S. pombe sporulation-specific coiled-coil protein spo15p is localized to the spindle pole body and essential for its modification."

RL J. Cell Sci. 113:545-554(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feitwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares S., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe."

RL Nature 415:871-880(2002).

RN [3]

RP SEQUENCE OF 705-871 FROM N.A.

RC STRAIN=968 h90;

RX MEDLINE=20223869; PubMed=10759899;

RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,

RA Hiraoaka Y.;

RT "Large-scale screening of intracellular protein localization in living fission yeast calls by the use of a GFP-fusion genomic DNA library."

RL Genes/Cells 5:169-190(2000).

CC -!- FUNCTION: Has a role in the initiation of spore membrane formation.

CC -!- SUBUNIT: Monomer.

CC -!- SUBCELLULAR LOCATION: Spindle pole body.

CC -!- SIMILARITY: Belongs to the MPC70 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z70690; CA94624.1; -.

DR EMBL; AB027811; BAA87115.1; -.

DR PIR; T38077; T38077.

KW GeneDB SPombe; SPAC1F3.06C; -.

KW Sporulation; Coiled coil.

FT DOMAIN 199 785 COILED COIL (POTENTIAL).

FT DOMAIN 804 1235 COILED COIL (POTENTIAL).

FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).

FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).

SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA0617D9DA CRC64;

Query Match 3.7%; Score 152; DB 1; Length 1957;

Best Local Similarity 18.5%; Pred. No. 0.22;

Matches 154; Conservative 146; Mismatches 313; Indels 220; Gaps 35;

Qy 25 TLADLIEHIEDHIDTPVLEKQE-----LQPTVVALSYINFMPTDAAR--R 71

Db 833 TLIDNVQKLMHVKVQSKVELKGVNKLSDLKRLNLSNVAISNDQILQLAELSK 892

Qy 72 EQSLKKKIQPKLSLTLSVSRGVSTPPRHSG-----SLTPPVTPPTPSSSR 123

Db 893 NYDSLEQ-----SAQLNSGLKSLAEKQLLHTEENELHRLDKLTOKLIEKSKSDLG 947

Qy 124 SSTPTPTQYRFLRTRNLIAPIFLHRTLTVMHSRNSRTNIKEKTFKVDMDLKVEMKMG 183

Db 948 KLTARQEELSN-LKEEN-----MSQQAITSVKS--KLDETLSKSKLEA 990

Qy 184 EQES-HLSAHLQITFTGPFHKNDKPSNSEN-ONSVTLEVLVVKVCHKR---KDVSC 238

Db 991 DIEHLKMKVSEVERNALIASNERLMDLLKNNGENIASLQTEI---EKKRAENDDLQS 1046

Qy 239 PIQVPTGKQVPLIPD-----LNQTK--PGNFPPLAVSSNRFEPNSHMKVSYSL 287

Db 1047 KLSVSVSEYENLLLISSQTKSLEDKTNQLKYEKKVQKLLDEKQNVLEELTSKYK 1106

Qy 288 LFRVTRPGRREFNGMINGETNENIDVNEELPARKENREDGE--KTFVAQMTVFDKNNRL 345

Db 1107 L-----GE--ENAIKDELLALRKSKSKQHDLCANFVDDLK--EKSDAL 1146

Qy 346 QLLDGEVEVAQWBEBCPISKATWETILDGKRLPPFTFSQGTPLQTLRTWGTENDK 405

Db 1147 EQLTNEKNEVLVSLEQ-----SINSNE 1168

Qy 406 STAPIAKPLATRNSE---SLHQENKPGSVKPTQIAVKESLTTLDTQTRKEKDTFENRQK 462

Db 1169 ALVEERSDLANRLSDMKSLSDSDNVISVIRSLVRVNDLDT--LKKDKDS----- 1218

Qy 463 LRIFYQLYNNTRQQTTEARDLHCPWCTLNCRKLYSLKHLKCHSRFTFNVYHPK-- 520

Db 1219 -----LSTQYSEVCQDRDD-----LDSLKGCBSF-NKYAVSLREL 1254

Qy 521 --GARIDVSNICYDGSY---AGNPDIHR-----QPQF----- 549



QY 622 DEKPEWLEKRTITQIEEFDVNEGEKVMKLNWLVFKH-----GFIADNMNHA 672  
 Db 566 -----YKLTSTIPKFDLQWGHQDAEAEFLTHLLDQLHEELISAIDGLTNE----- 612  
 QY 673 CMLFVENVYQKLIKLNCLNENMLHLVSMH---DENLISIMSDKAVTKLRENOCKLEKGE 729  
 Db 613 ----IQMQLQSNDEQL-KVFFIRMLSRYGKAEF-----IKNASPRIKELIEKY---- 656  
 QY 730 SASPANEETERQNGTANGFSEINSKEKALETDSVGSVKQSKKOK 775  
 Db 657 --GVINDDSTEB-----NGWHE-----VSGSSKRGKTK 683

# RESULT 14

USOL\_YEAST STANDARD; PRT; 1790 AA.  
 AC P25386;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Intracellular protein transport protein USOL1.  
 GN USOL1 OR INT1 OR YDL058W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=X2180-1A;  
 RX MEDLINE=91185402; PubMed=2010462;  
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,  
 RA Yamasaki M.;  
 RT "A cytoskeleton-related gene, usol1, is required for intracellular  
 protein transport in Saccharomyces cerevisiae.";  
 RL J. Cell Biol. 113:245-260(1991).  
 RN [2]  
 RP SEQUENCE OF 782-1790 FROM N.A.  
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,  
 RA Kendrick K.E.;  
 RL Submitted (FEB-1993) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RA Bai Y., Symington L.S.;  
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.  
 CC -!- FUNCTION: Required for protein transport from the ER to the Golgi  
 complex.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR  
 MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE  
 ER AND THE GOLGI COMPLEX.  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed  
 of an heptapeptide repeat pattern characteristic of alpha-helical  
 coiled coils. May form filamentous structures in the cell.  
 CC -!- SIMILARITY: BELONGS TO THE VDP/USOL1/YBL047C FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X54378; CAA38253.1; -;  
 DR EMBL; L03188; AAB00143.1; -;  
 DR EMBL; U53668; AAB66559.1; -;  
 DR GeneOnline; 140300; -;  
 DR SGD; S0002216; USOL1.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR002017; Spectrin.  
 DR InterPro; IPR006955; Usol\_p115\_C.  
 DR InterPro; IPR006953; Usol\_p115\_head.  
 DR Pfam; PF04871; Usol\_p115\_C\_1.  
 DR Pfam; PF04869; Usol\_p115\_head; 1.

Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.  
 KW DOMAIN 1 724 GLOBULAR HEAD.  
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).  
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).  
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.  
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).  
 FT CONFLICT 847 847 G -> E (IN REF. 2).  
 FT CONFLICT 924 924 E -> K (IN REF. 2).  
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).  
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).  
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).  
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).  
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).  
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).  
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).  
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216B9FD4818 CRC64;  
 Query Match 3.68; Score 147.5; DB 1; Length 1790;  
 Best Local Similarity 18.7%; Pred. No. 0.37;  
 Matches 161; Conservative 139; Mismatches 351; Indels 211; Gaps 34;  
 QY 29 LIEHIEDNHIDTP-RVLEK--QELQPTVALSYINRMTD--AARREOESLKTKQPK 83  
 Db 827 LLEYKSTIHQEDSIKLEKLEGLTILLSQKKAEDGINKMGKDLFALSREMQAVEENCK-- 884  
 QY 84 LSLTSSSVSRGNVSTPPRHSSGSLTPPVTPPTTPSSSFRSSTPTPTQIYRFLRNLI 143  
 Db 885 ---NLQEKDKSNVN--HQETKSLKEDIAAKITEIKAINENLEEMKIQCNNLSKEKEHI 939  
 QY 144 APIFLHRTLTMYGHRNSRTNIKTEKVDMLSKVEKMGKEQS-----HLSAHLQLT 198  
 Db 940 SKELVEKSRPQSHDL---VAKTEKLKSLANNYKDMQAEENSLKAVEESKNESIQ 996  
 QY 199 TGFHFKNDKPSNSENQNSVTLEVLVVKVCHKRDVSCPIRQVPTGKQVPLIPLNQ 258  
 Db 997 SNLQNKIDMSQKENFQ-----IERGSIKKNIEQL---KK---TISLEQ 1036  
 QY 259 TKGNFPPLAVSSNEPE-----PNSHMVKSYSLLFRVTRGRREFNGMIN 304  
 Db 1037 TKBEIISKSSDKDEYSQISLKEKLETATTANDENVNKISBLTTRBELEAELAAKN 1096  
 QY 305 --GETNENIDVNBELPARRKRNRD--GEKTFVQMTVFDKRNRLQLIDGVEYVAMQEMEE 361  
 Db 1097 LKNELETKETSEKALKEVENEHLKEEIKLEKEATETKQQLNSLRANLESLEKEDH 1156  
 QY 362 CPISKKATWETILDGKLPPFTFGQPTLQTLRWGTNDKSTAPIAKPATRNSES 421  
 Db 1157 LAAQLKK--YEEQANKERQYNEISQ-----LNDEITS-----TQENES 1195  
 QY 422 LHQENK--PGSVKPTOTIAVKS-----LTTDLQTEKEKDTNENR--OKLRIFYQ 468  
 Db 1196 IKKNDLEGEVAMKSTSEQSNLKKSEIDALNLQIKELKKKNETNEASLLSEIK--- 1251  
 QY 469 FLYNNTRQOTEARDDLHCPWCTLNCRKLYSLKLHLKCHSRFIFNYVHPKGRIDVSI 528  
 Db 1252 -SVESETVKIKELQDE-----CNFKEKVESELEDKLKASEDK-NSKYLELQKSE--KI 1301  
 QY 529 NECVDGSGYAGNPQDIHQPGFASRNGPVKZETITHLVCRPKTKASSEFSEDEGEV 588  
 Db 1302 KEELDAK-----TTTELKLEKIT--NLSKAKEKSEEL 1333  
 QY 589 EQORTYSSGHNRLYFHSDTCLPLRPOEMVEDSD-EKDPWLREKTTITQIEEFD-VNEG 646  
 Db 1334 SRLKKTSSERK--NAEEQLEKLNKEIKQNAFAEKERKLLNEGSSITITQIESEKINTL 1390  
 QY 647 EKEVMKLNWLVHMKHGFADNQNHACMLFVENYQKIKKLNCRNFMHLVS-----M 700  
 Db 1391 EDELIRQENELKAKEI-DNTRSELEKVSLSNDELLEEKQNTIKSLQDELILSKDKITR 1449  
 QY 701 HDEFNLISI-----MSIDKA-VTKLREMOOK 724  
 Db 1450 NDEKLLSIEDNKEDLESLEQLRAAQESKAKVEEGLKLEBESSKEKAELEKSKEMWKK 1509

QY 725 LEK-----GASAPANEITEEQNGTANGFSEINSKEKALE- 760  
 DB 1510 LESTIESNETELKSMETIRKSDLEQSKASABEDIKNLQHEKSDLSIRNESEKIDIE 1569  
 QY 761 -----TDSVGVSKOSKQKL 776  
 DB 1570 LKSKLRIEAKSGSELETVKQEL 1591

RESULT 15  
 ID GOA4 HUMAN STANDARD; PRT: 2230 AA.  
 AC Q13439; Q13270; Q13654; Q14436;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE GOlgi autoantigen, golgin subfamily A member 4 (Trans-Golgi p230) (256  
 DE xDa golgin) (Golgin-245) (72.1 protein).  
 GN GOLGA4.  
 OS Homo sapiens (Human)  
 OC Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=96215236; PubMed=8626529;  
 RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;  
 RT "Molecular characterization of trans-Golgi p230: a human peripheral  
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains  
 RT extensive coiled-coil alpha-helical domains and a granin motif.";  
 RL J. Biol. Chem. 271:8328-8337(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Seelig H.P.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 131-2230 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan B.K.L.;  
 RT "Molecular characterization of golgin-245, a novel Golgi complex  
 RT protein containing a granin signature.";  
 RL J. Biol. Chem. 270:31262-31268(1995).  
 RN [4]  
 RP SEQUENCE OF 524-672 FROM N.A.  
 RC TISSUE=Gastric fundus;  
 RA Balague C.;  
 RL Thesis (1994), Instituto municipal de investigacion medica, Spain.  
 CC -!- FUNCTION: May play a role in vesicular transport from the trans-  
 CC Golgi.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein  
 CC associated with the trans-Golgi network.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q13439-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q13439-2; Sequence=VSP\_004272, VSP\_004273;  
 CC Name=3;  
 CC IsoId=Q13439-3; Sequence=VSP\_004274;  
 CC Name=4;  
 CC IsoId=Q13439-4; Sequence=VSP\_004275;  
 CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.  
 CC -!- DISEASE: Antibodies against GOLGA4 are present in sera from  
 CC patients with Sjogren's syndrome (SS) [MIM:270150]. Sera from  
 CC patients with SS often contain antibodies that react with normal  
 CC components of the Golgi complex.  
 CC -!- DISEASE: Antibodies against GOLGA4 are found in sera from  
 CC hepatitis B patients.  
 CC -!- SIMILARITY: Belongs to the golgin family.  
 CC -!- SIMILARITY: Contains 1 GRIP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U41740; AAC50434.1; -;  
 DR EMBL; X82834; CAA58041.1; -;  
 DR EMBL; U31906; AAC51791.1; -;  
 DR EMBL; X76942; CAA54261.1; -;  
 DR Genew; HGNC:4427; GOLGA4.  
 DR MIM; 602509; -;  
 DR MIM; 270150; -;  
 DR GO; GO:0003802; C:Golgi trans face; TAS.  
 DR GO; GO:0016192; P:vesicle-mediated transport; TAS.  
 DR InterPro; IPR000237; GRIP\_domain.  
 DR Pfam; PF01465; GRIP; 1.  
 DR PROSITE; PS50913; GRIP; 1.  
 KW Antigen; Golgi stack; Coiled coil; Alternative splicing.  
 FT DOMAIN 133 2185  
 FT DOMAIN 252 2096  
 FT DOMAIN 2168 2215  
 FT VARSPLIC 2154 2185  
 FT AICTRMGSHSLNPLNLSICEKCLSI (in isoform  
 FT 2).  
 FT /FTId=VSP\_004272.  
 FT Missing (in isoform 2).  
 FT /FTId=VSP\_004273.  
 FT Missing (in isoform 3).  
 FT /FTId=VSP\_004274.  
 FT /FTId=VSP\_004275.  
 FT /FTId=VSP\_004275.  
 FT R -> K (IN REF. 3).  
 FT Y -> H (IN REF. 3).  
 FT T -> A (IN REF. 3).  
 FT K -> E (IN REF. 3).  
 FT T -> A (IN REF. 3).  
 FT K -> E (IN REF. 3).  
 FT K -> E (IN REF. 3).  
 FT K -> N (IN REF. 3).  
 FT CONFLICT 682 682  
 SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;  
 Query Match 3.6%; Score 147; DB 1; Length 2230;  
 Best Local Similarity 18.2%; Pred. No. 0.53;  
 Matches 152; Conservative 148; Mismatches 308; Indels 228; Gaps 36;  
 QY 30 IEHIEDNHIDTPRVLEK-----QELQOPTYVALSVINREMTDAARQESLKKIQ 81  
 DB 330 LEKIKDLHMAEKTLITQLRDKNLIEQLEQKGVIAETKQMHETLEMEKEEIAQLRS 389  
 QY 82 PKLSLTLSSSVSRGNVSTPPRHSSGLTPPVTPPTPSSPSFSSPTPTPTIYRFLTRN 141  
 DB 390 RIKQMTTQGEELREQEKESRAAFBELEKALSTAQKTEARKLKAEMDEQIKTEKTS 449  
 QY 142 LIAPFLHRLTYMSHRNSRTNIKTEKV-----DDMLSKVEKM-----KGEQESHSL 190  
 DB 450 -----EERISLQQLSRVKQEVVDVMKKSSEEQIALQKLHEKELARKEQ-----L 496  
 QY 191 SAHLQTLTFTGFHKNDKPSNSENPQNSVTLEVL---LVKVCCHKRKDVCFIRQVPTGK 247  
 DB 497 TKKLQTRREF-----QEQMVALEKSQSEYVKLSQBEKQEQESLALBELELQK 544  
 QY 248 KOV-----PLIPDLNQTQPGNPPSLAVSNPEPNSHMVKSYSLLFVTPGPREFGM 302  
 DB 545 KAILTESENKURDLQ-----EAETRYTRILELESLSKSLQENKQSKDL 590  
 QY 303 ---INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRQLQLDGEYEVAMQEM 359  
 DB 591 AVHEAEKXKH---NKEITVAVVEKHTELESIAKHQODALW---TEKLOVLKQOYQTEMEKL 645  
 QY 360 -BECPISKRATWETILDGKLLPPTFSQPTLQFTLRWTGETNDSKTAPI-----AK 412

```

646 REKCEQEK-----ETLLKDKEI-IPAHE-----ENNEKTEKLDVQKTELE 687
413 PLATRNSESL---HOENKPGSVKPTQTIAVKESSLTDLQTRKEKDPNENRQKLRIFYQF 469
688 SLSELSSEVLKARHKLEELSVLKQDQDKMQELEAKWDEQK-----NHHQQQVDSIIKE 742
470 LYNNTRQOTEARDDLHCPWCVTILNCRKYSLL---KHLKLCHSRFIFNVYVHPKGARD 525
743 HEVSIQRTKALKQDIN-----QUELLKSRDKHLK-----BHQAHVENLEAD 785
526 VSNIECYDGSYAGNQDIIHQPGFAFSRNGPVKPTITHILCRPKETKASMEF-----580
786 IKKSE-GELOQASAKLDVFQSYSA-----TH-----FQTKAYEOLLAQLOQ 826
581 -----JESD-----GEVEQORTYSVSGHNLRYPHSDTCLPLRPQMEVDS-----EDEK 624
827 KLULD-ETERILLTKQVAEYEQK-----KDVCTELDAHKIQVLDLMQOLEKQN 874
625 DEPWLRKKTITQIEFFSDVNEGEKEVMKLNHLHYMKHGFADNQMHACHMLFVENYGOKI 684
875 SEMEQKVSLLTOVYE-SKLEDGNKEQEQTKQILVEKENMIL--OMEGQKKEIILTKQL 931
685 IKKNLCRNFMHLVSHWHPNLISIMSDIKAVTKLRMEMQKLEKGEASAPANEE-----737
932 SAKK-----DSIHILN-----ZEYETKEFQKQKWKVKQAKEMQETLLKKLL 975
738 -----ITHEQNTANGFSEINSKEKALE-----TDSVSGVSGVSKSQKK 774
976 DOEAKLKELENTA---LELSQKEQFNKAKLEMAQANSAGISDAVSRLTNQKEQ 1028

```

Search completed: August 25, 2004, 17:55:08  
Job time : 33 secs







Query Match	83.4%; Score 647; DB 3; Length 803;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 647; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	130 PTQIYRFLRTRNLIAPIFLHRTLTYSNHRNRTNKKRTFKVDDMLSKVKEMKGEQESH 189
DB	157 PTQIYRFLRTRNLIAPIFLHRTLTYSNHRNRTNKKRTFKVDDMLSKVKEMKGEQESH 216
QY	130 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRQVSCPIRQVPTGKKQ 249
DB	217 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRQVSCPIRQVPTGKKQ 276
QY	250 VPLIPDLNQTGKPNFPPSLAVSSNEFEPSPNSHMKVSYLLFRVTRPGRREFNGMGETNE 309
DB	277 VPLIPDLNQTGKPNFPPSLAVSSNEFEPSPNSHMKVSYLLFRVTRPGRREFNGMGETNE 336
QY	310 NIDVNEELPARKKRNREDGKTFVAQMTVPDKNRRLQLLDGEYEVAMQEMEECPISKRA 369
DB	337 NIDVNEELPARKKRNREDGKTFVAQMTVPDKNRRLQLLDGEYEVAMQEMEECPISKRA 396
QY	370 TWETILDGKRLPPPTFTFSOGPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKPG 429
DB	397 TWETILDGKRLPPPTFTFSOGPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKPG 456
QY	430 SVKPTQTIAVRESLTDLQTRKEDTNPENRQKLRIFVQFLYNNNTROQTEARDLHCPW 489
DB	457 SVKPTQTIAVRESLTDLQTRKEDTNPENRQKLRIFVQFLYNNNTROQTEARDLHCPW 516
QY	490 CTLNCRKLYSLKHLKCHSRFIFNYVHPKGRIDVSNICYDGSVAGNPQDIHRQPGF 549
DB	517 CTLNCRKLYSLKHLKCHSRFIFNYVHPKGRIDVSNICYDGSVAGNPQDIHRQPGF 576
QY	550 AFSRNGPVKRPITPILVCRPKTKASREPLESEDEGEVQORYSSGHNLVYFHSDTCL 609
DB	577 AFSRNGPVKRPITPILVCRPKTKASREPLESEDEGEVQORYSSGHNLVYFHSDTCL 636
QY	610 PLRQEMEVDSDEKDPWLREKTIQIEEFSDVNEGEKVMKLNHLVNMKGHFIADNQ 669
DB	637 PLRQEMEVDSDEKDPWLREKTIQIEEFSDVNEGEKVMKLNHLVNMKGHFIADNQ 696
QY	670 NHACMLFVNTGQKIIKQVCRNFMHLVSHMDNLSIMSIDKAVTKLREMOKLEKGE 729
DB	697 NHACMLFVNTGQKIIKQVCRNFMHLVSHMDNLSIMSIDKAVTKLREMOKLEKGE 756
QY	730 SASPANEIITEBQNTANGFSEINSKEKALETDSVGSVQSKQKOL 776
DB	757 SASPANEIITEBQNTANGFSEINSKEKALETDSVGSVQSKQKOL 803
RESULT 2	
ID	ABG30842
XX	ABG30842 standard; protein; 739 AA.
AC	ABG30842;
XX	
DT	21-OCT-2002 (first entry)
DE	Human joined with jAZF1 (jAZ1) protein.
KW	Human; jAZF1; juxtaposed with another zinc finger; jAZ1; jAZF1/jAZ1;
KW	joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;
KW	antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
KW	chromosome 17.
OS	Homo sapiens.
XX	
FH	Key
FT	Region
FT	Location/Qualifiers
FT	92...93
FT	/note= "Breakpoint for production of fusion protein
FT	jAZF1/jAZ1"
FT	448...471
FT	/note= "Encoded by CTGCATTGCCCTTGCTGTACTCTGAAGTCGGCGCAA-
FT	ACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT. C2H2 zinc finger

FT	Region	domain"
FT	521...538	
XX	/note= "Bipartite nuclear localisation signal"	
XX	WO200193805-A2..	
XX	13-DEC-2001..	
XX	04-JUN-2001; 2001WO-US017936.	
XX	02-JUN-2000; 2000US-0209093P.	
XX	(BCEM ) BRIGHAM & WOMENS HOSPITAL INC.	
XX	Koontz J, Sklar J;	
XX	WPI; 2002-575047/61.	
XX	N-PSDB; ABK89162.	
XX	Novel jAZF1, jJAZ1 or jAZF1/jJAZ1 polypeptides useful as immunogens or	
XX	antigens to raise or test anti-jAZF1, jJAZ1 or jAZF1/jJAZ1 antibodies.	
XX	Claim 5; Fig 2; 76pp; English.	
XX	The present invention relates to a new jAZF1 (juxtaposed with another	
XX	zinc finger), jJAZ1 (joined with jAZF1) or jAZF1/jJAZ1 polypeptide. The	
XX	methods of the invention can be used to identify a compound which	
XX	controls proliferation of endometrial stroma, by expressing jJAZ1 in the	
XX	presence of the compound, and determining whether the compound affects	
XX	expression of jJAZ1, jJAZ1 or jAZF1/jJAZ1 polypeptides are useful	
XX	as immunogens or antigens to raise or test anti-jAZF1, jJAZ1 or	
XX	jAZF1/jJAZ1 antibodies. The invention can be used as bait proteins in a	
XX	two hybrid assay or three hybrid assay to identify other proteins which	
XX	bind or interact with jAZF1/jJAZ1-binding proteins. jAZF1, jJAZ1 or	
XX	jAZF1/jJAZ1 molecules are useful for identifying the origin of tumour and	
XX	as tumour marker protein to verify that a stromal tumour is from	
XX	endometrium. The antibody is useful for promoting or decreasing fertility	
XX	or pregnancy, and also for treating endometrial stromal tumours. The	
XX	present amino acid sequence represents the human jJAZ1 protein of the	
XX	invention. This sequence is encoded by the human jJAZ1 gene located on	
XX	chromosome 17	
XX	Sequence 739 AA;	
XX	Query Match 45.7%; Score 355; DB 5; Length 739;	
XX	Best Local Similarity 100.0%; Pred. No. 0;	
XX	Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	130 PTQIYRFLRTRNLIAPIFLHRTLTYSNHRNRTNKKRTFKVDDMLSKVKEMKGEQESH 189	
DB	93 PTQIYRFLRTRNLIAPIFLHRTLTYSNHRNRTNKKRTFKVDDMLSKVKEMKGEQESH 152	
QY	190 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRQVSCPIRQVPTGKKQ 249	
DB	153 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRQVSCPIRQVPTGKKQ 212	
QY	250 VPLIPDLNQTGKPNFPPSLAVSSNEFEPSPNSHMKVSYLLFRVTRPGRREFNGMGETNE 309	
DB	213 VPLIPDLNQTGKPNFPPSLAVSSNEFEPSPNSHMKVSYLLFRVTRPGRREFNGMGETNE 272	
QY	310 NIDVNEELPARKKRNREDGKTFVAQMTVPDKNRRLQLLDGEYEVAMQEMEECPISKRA 369	
DB	273 NIDVNEELPARKKRNREDGKTFVAQMTVPDKNRRLQLLDGEYEVAMQEMEECPISKRA 332	
QY	370 TWETILDGKRLPPPTFTFSOGPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKPG 429	
DB	333 TWETILDGKRLPPPTFTFSOGPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKPG 392	
QY	430 SVKPTQTIAVRESLTDLQTRKEDTNPENRQKLRIFVQFLYNNNTROQTEARD 484	
DB	393 SVKPTQTIAVRESLTDLQTRKEDTNPENRQKLRIFVQFLYNNNTROQTEARD 447	

RESULT 3  
 ABG30843  
 ID ABG30843 standard; protein; 776 AA.  
 XX  
 AC  
 AC ABG30843;  
 XX  
 21-OCT-2002 (first entry)  
 XX  
 DE Human JAZF1/jJAZ1 fusion protein.  
 DE  
 DE  
 XX  
 KW Human; JAZF1; juxtaposed with another zinc finger; jJAZ1; JAZF1/jJAZ1;  
 KW joined with JAZF1; proliferation; endometrial stroma tumour; immunogen;  
 KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;  
 KW fusion protein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS Chimeric.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 13...37  
 FT /note= "Encoded by ACCTGCCGATCGGGGGCTGGCGACTCCATTC-  
 FT CCCACCTGGCGGACCTCATCGACGACGAGGACCAACAC. C2H2 zinc  
 FT finger domain"  
 FT 99...100  
 FT /note= "Breakpoint for production of fusion protein  
 FT JAZF1/jJAZ1"  
 FT  
 FT Misc-difference 357  
 FT /note= "Encoded by CAG"  
 FT  
 FT Misc-difference 485...508  
 FT /note= "Encoded by CTGCATTGCTTGGTGTTACTCTGAACTCCGCAA-  
 FT ACTTATAGTTTACTCAAGCATCTTAACTTGCCAT. C2H2 zinc finger  
 FT domain"  
 FT 558...575  
 FT /note= "Bipartite nuclear localisation signal"  
 FT  
 XX  
 PN WO200193805-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 XX 04-JUN-2001; 2001WO-US017936.  
 XX  
 XX 02-JUN-2000; 2000US-0209093P.  
 XX  
 XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
 XX  
 XX Koontz J, Sklar J;  
 XX  
 DR WPI; 2002-575047/61.  
 DR N-PSDB; ABR89163.  
 XX  
 XX Novel JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides useful as immunogens or  
 PT antigens to raise or test anti-JAZF1, jJAZ1 or JAZF1/jJAZ1 antibodies.  
 PT  
 XX  
 PS Claim 7; Fig 3; 76pp; English.  
 XX  
 CC The present invention relates to a new JAZF1 (juxtaposed with another  
 CC zinc finger), jJAZ1 (joined with JAZF1) or JAZF1/jJAZ1 polypeptide. The  
 CC methods of the invention can be used to identify a compound which  
 CC controls proliferation of endometrial stroma, by expressing jJAZ1 in the  
 CC presence of the compound, and determining whether the compound affects  
 CC expression of jJAZ1, jJAZ1 or JAZF1/jJAZ1 polypeptides are useful  
 CC as immunogens or antigens to raise or test anti-JAZF1, jJAZ1 or  
 CC JAZF1/jJAZ1 antibodies. The invention can be used as bait proteins in a  
 CC two hybrid assay or three hybrid assay to identify other proteins which  
 CC bind or interact with JAZF1/jJAZ1-binding proteins. JAZF1, jJAZ1 or  
 CC JAZF1/jJAZ1 molecules are useful for identifying the origin of tumour and  
 CC as tumour marker protein to verify that a stromal tumour is from  
 CC endometrium. The antibody is useful for promoting or decreasing fertility  
 CC or pregnancy, and also for treating endometrial stromal tumours. The  
 CC present amino acid sequence represents the human JAZF1/jJAZ1 fusion  
 CC protein of the invention

SQ Sequence 776 AA;  
 Query Match 44.6%; Score 346; DB 5; Length 776;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 38 IDTDPVLEKQELQOQPTVALSYINRPMTPDAARREESLKKIOPKLSLTSSSVSGNV 97  
 DB 38 IDTDPVLEKQELQOQPTVALSYINRPMTPDAARREESLKKIOPKLSLTSSSVSGNV 97  
 QY 98 STPRHSSGSLTPVTPPTIPSSSFRSSTPTPTQIYRFLTRNLIAPIFLHRTLTVMH 157  
 DB 98 STPRHSSGSLTPVTPPTIPSSSFRSSTPTPTQIYRFLTRNLIAPIFLHRTLTVMH 157  
 QY 158 RNSRTNKRKTFKVDMLSKVKMGQESHSLSAHLQLTFTGFFHKNDKPSNSENEQN 217  
 DB 158 RNSRTNKRKTFKVDMLSKVKMGQESHSLSAHLQLTFTGFFHKNDKPSNSENEQN 217  
 QY 218 SVTLLEVLVVKCHKKRQVSCPIROVPTGKKQVPLPDLNCTKGNFPPSLAVSSNEPEPS 277  
 DB 218 SVTLLEVLVVKCHKKRQVSCPIROVPTGKKQVPLPDLNCTKGNFPPSLAVSSNEPEPS 277  
 QY 278 NSHVKVSYLLFVTRPGRREFNMGINGETNENIDVNEELPARRKRNREDGKTFVAQMT 337  
 DB 278 NSHVKVSYLLFVTRPGRREFNMGINGETNENIDVNEELPARRKRNREDGKTFVAQMT 337  
 QY 338 VFDKRRLOLLDGEYEVAMQEMEECPISKRAWEITLDGKRLPPPTPSQGTLOFTLR 397  
 DB 338 VFDKRRLOLLDGEYEVAMQEMEECPISKRAWEITLDGKRLPPPTPSQGTLOFTLR 397  
 QY 398 WTGETNDKSTAPAKPLATRNSESLHOENKPGSVKPTQTIIVKESLTITDLOTRKEKDTN 457  
 DB 398 WTGETNDKSTAPAKPLATRNSESLHOENKPGSVKPTQTIIVKESLTITDLOTRKEKDTN 457  
 QY 458 ENRQKLRFYQFLYNNNTROQTEARDD 484  
 DB 458 ENRQKLRFYQFLYNNNTROQTEARDD 484  
 RESULT 4  
 AAU15958  
 ID AAU15958 standard; protein; 289 AA.  
 XX  
 AC AAU15958;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 XX Human novel secreted protein, Seq ID 911.  
 DE  
 DE  
 XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
 KW antibacterial; virucide; fungicide; ophthalmological; vulnary; secreted  
 KW protein; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
 KW preservative; antiproliferative.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200155322-A2.  
 PN  
 XX  
 XX 02-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US001341.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 30-AUG-2000; 2000US-0227009P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 02-OCT-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 13-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250191P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0251990P.  
 PR 05-JAN-2001; 2000US-0254097P.  
 PR 05-JAN-2001; 2000US-0259678P.

(HUKA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-488783/53.

N-PSDB; AAS25945.

New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives.

Claim 11; SEQ ID NO 911; 980pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Antibodies to the proteins can also be used in  
CC alleviating symptoms associated with the disorders and in diagnostic  
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
CC and many other disorders listed in the specification. The polypeptides  
CC can also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention. Note: The  
CC sequence data for this patent did not form part of the printed

Query Match 36.5%; Score 283; DB 4; Length 289;  
Best Local Similarity 100.0%; Pred. No. 6.9e-277; Indels 0; Gaps 0;  
Matches 283; Conservative 0; Mismatches 0;  
QY 494 CRKLSLLKHLKLSRIFVYVHPKGRIDVSNICVDSYAGNPQDIHQPCGFAPSR 553  
Db 7 CRKLSLLKHLKLSRIFVYVHPKGRIDVSNICVDSYAGNPQDIHQPCGFAPSR 66  
QY 554 NQPVKRTITHLVCRPKTKASMEFLESDGEVEQRTSSGHNRLYFSDTCLPLRP 613  
Db 67 NQPVKRTITHLVCRPKTKASMEFLESDGEVEQRTSSGHNRLYFSDTCLPLRP 126  
QY 614 QMEVDSDEKDPWLRKTTIQIEFSDVNEGEKVMKLNHLVNMKGFIADQMNHAC 673  
Db 127 QMEVDSDEKDPWLRKTTIQIEFSDVNEGEKVMKLNHLVNMKGFIADQMNHAC 186  
QY 674 MLEFVNYGQKIKNLCRNFMHLVSMHDFNLISIMSDKAVTKLREMQQKLEKGEASGP 733  
Db 187 MLEFVNYGQKIKNLCRNFMHLVSMHDFNLISIMSDKAVTKLREMQQKLEKGEASGP 246  
QY 734 ANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
Db 247 ANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 289

RESULT 5  
ABU55027  
ID ABU55027 standard; protein; 289 AA.  
XX AC ABU55027;  
XX DT 18-MAR-2003 (first entry)  
XX DE Human novel polypeptide #114.  
XX KW Human; neural disorder; immune system disorder; renal disorder;  
KW muscular disorder; respiratory disease; reproductive disorder;  
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
KW haemostatic; antiarteriosclerotic.  
XX OS Homo sapiens.  
XX PN US2002132753-A1.  
XX PD 19-SEP-2002.  
XX PF 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180828P.  
PR 28-JUN-2000; 2000US-0214986P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 22-AUG-2000; 2000US-0225758P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234597P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.

(ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-147444/14.  
XX N-PSDB; ABX73286.  
XX New polypeptides and nucleic acids, useful in gene therapy for treating,  
PT inhibiting or preventing e.g. neural, immune system, muscular,  
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
PT renal disorders.  
XX Claim 11; SEQ ID NO 911; 402pp; English.  
XX The invention relates to human novel polypeptides and their associated  
XX polynucleotides. The polypeptides and polynucleotides are useful in gene  
CC therapy for treating, inhibiting or preventing neural disorders, immune  
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.

CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
 CC ABUS5748 represent human novel polypeptides of the invention  
 XX  
 SQ Sequence 289 AA;

Query Match 36.5%; Score 283; DB 6; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-277;  
 Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 494 CRKLYSLKHLKCHSRIFNYVHPKGRIDVINSICYDGSYAGNPQDIHQPGAFSR 553  
 DB 7 CRKLYSLKHLKCHSRIFNYVHPKGRIDVINSICYDGSYAGNPQDIHQPGAFSR 66  
 QY 554 NGPVKRTPIHILVCRPKRTKASMSFLESEDEGEQRTYSSGHNRLYFHSDTCLPLRP 613  
 DB 67 NGPVKRTPIHILVCRPKRTKASMSFLESEDEGEQRTYSSGHNRLYFHSDTCLPLRP 126  
 QY 614 QMEVDSDEKDPWLREKTTIQLEESDVNVEGEKVMKLNHVMKHGFIADNQMHAC 673  
 DB 127 QMEVDSDEKDPWLREKTTIQLEESDVNVEGEKVMKLNHVMKHGFIADNQMHAC 186  
 QY 674 MLFVNYQCKIIKKNLCRNFMLHLSMDFNLISIMSDKAVTKLRMQOKLEKGEASGP 733  
 DB 187 MLFVNYQCKIIKKNLCRNFMLHLSMDFNLISIMSDKAVTKLRMQOKLEKGEASGP 246  
 QY 734 ANEETEEQNTANGFSEINSKEKALETDSVSGVSKOSKKQKL 776  
 DB 247 ANEETEEQNTANGFSEINSKEKALETDSVSGVSKOSKKQKL 289

RESULT 6

AAU15978  
 ID AAU15978 standard; protein; 388 AA.  
 AC AAU15978;  
 XX  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human novel secreted protein, Seq ID 931.  
 XX  
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 KW cardiac; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
 KW preservative; antiproliferative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155322-A2.  
 XX  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001341.  
 XX  
 XX 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR

PR 19-MAY-2000; 2000US-0205515P.  
 PR 28-JUN-2000; 2000US-0209467P.  
 PR 30-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 19-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0228287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 21-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR

PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(PUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-488783/53.  
XX N-PSDB; AAS25965.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
XX diagnosing, preventing, treating or ameliorating medical conditions and  
XX used as food additives or preservatives.  
XX  
XX Claim 11; SEQ ID NO 931; 980pp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their  
XX encoded secreted proteins. The nucleic acids and proteins are used to  
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
XX

CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Antibodies to the proteins can also be used in  
CC alleviating symptoms associated with the disorders and in diagnostic  
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays  
CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,  
CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
CC and many other disorders listed in the specification. The polypeptides  
CC can also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention. Note: The  
CC sequence data for this patent did not form part of the printed

Query Match 24.0%; Score 186; DB 4; Length 388;  
Best Local Similarity 99.5%; Pred. No. 1e-178;  
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 389 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQNKGSGVKPTQTIAVKESLTDLQ 448  
Db 1 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQNKGSGVKPTQTIAVKESLTDLQ 60  
Qy 449 TRKEKTPNENRQKLRIFYQFLYNNTRQTEARDLHCPWCTLNCRKLYSLKHLKCH 508  
Db 61 TRKEKTPNENRQKLRIFYQFLYNNTRQTEARDLHCPWCTLNCRKLYSLKHLKCH 120  
Qy 509 SRPIFNVYHPKGARIDVSNIECYGAGNPDIIHQPGFASRNGPVKRTPIITHLVC 568  
Db 121 SRPIFNVYHPKGARIDVSNIECYGAGNPDIIHQPGFASRNGPVKRTPIITHLVC 180  
Qy 569 RPKRTKASMEFLESEDEGEVEQRTYSSGHNRLYFHSDTCLPLRQMEVDESDKDEW 628  
Db 181 RPKRTKASMEFLESEDEGEVEQRTYSSGHNRLYFHSDTCLPLRQMEVDESDKDEW 240  
Qy 629 LREKTIQIEFSDVNEGEKEVMKLNHLVHMKGFTADNMHACMLFVENYQKTIKKN 688  
Db 241 LREKTIQIEFSDVNEGEKEVMKLNHLVHMKGFTADNMHACMLFVENYQKTIKKN 300  
Qy 689 LCRNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKESASPANEETEPQNGTANG 748  
Db 301 LCRNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKESASPANEETEPQNGTANG 360  
Qy 749 FSEINSKEKALETDSVSGVSKQSKOKL 776  
Db 361 FSEINSKEKALETDSVSGVSKQSKOKL 388

RESULT 7  
ABU55047  
ID ABU55047 standard; protein; 388 AA.  
XX  
XX AC ABU55047;  
XX  
XX DT 18-MAR-2003 (first entry)  
XX  
XX DE Human novel polypeptide #134.  
XX  
XX Human; neural disorder; immune system disorder; renal disorder;  
XX muscular disorder; respiratory disease; reproductive disorder;  
XX gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
XX hyperproliferative disorder; inflammatory disease; allergic reaction;  
XX blood related disorder; cancer; immunosuppressive; antiinflammatory;  
XX cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
XX haemostatic; antiarteriosclerotic.  
XX



OS	Homo sapiens.	PS	Claim 11; SEQ ID NO 931; 402pp; English.
XX		XX	The invention relates to human novel polypeptides and their associated
PR	US2002132753-A1.	CC	polynucleotides. The polypeptides and polynucleotides are useful in gene
XX		CC	therapy for treating, inhibiting or preventing neural disorders, immune
PD	19-SEP-2002.	CC	system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
XX		CC	and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
PF	17-JAN-2001; 2001US-00764864.	CC	nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
XX		CC	gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
PR	31-JAN-2000; 2000US-0179065P.	CC	(e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
PR	04-FEB-2000; 2000US-0180628P.	CC	heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
PR	28-JUN-2000; 2000US-0214896P.	CC	renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
PR	07-JUL-2000; 2000US-0216647P.	CC	leukemia), inflammatory diseases (e.g. septic shock, bursitis and
PR	07-JUL-2000; 2000US-0216880P.	CC	appendicitis), allergic reactions and conditions (e.g. asthma), blood
PR	11-JUL-2000; 2000US-0217487P.	CC	related disorders (e.g. thrombosis, atherosclerosis and myocardial
PR	11-JUL-2000; 2000US-0217496P.	CC	infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
PR	14-JUL-2000; 2000US-0218290P.	CC	ABUS5748 represent human novel polypeptides of the invention
PR	26-JUL-2000; 2000US-0220963P.	XX	
PR	26-JUL-2000; 2000US-0220964P.	SQ	Sequence 388 AA;
PR	14-AUG-2000; 2000US-0224518P.		
PR	14-AUG-2000; 2000US-0224519P.		
PR	14-AUG-2000; 2000US-0225267P.		
PR	14-AUG-2000; 2000US-0225268P.		
PR	14-AUG-2000; 2000US-0225270P.		
PR	14-AUG-2000; 2000US-0225447P.		
PR	14-AUG-2000; 2000US-0225757P.		
PR	14-AUG-2000; 2000US-0225758P.		
PR	22-AUG-2000; 2000US-0226888P.		
PR	30-AUG-2000; 2000US-0228924P.		
PR	01-SEP-2000; 2000US-0228928P.		
PR	01-SEP-2000; 2000US-0229343P.		
PR	01-SEP-2000; 2000US-0229344P.		
PR	01-SEP-2000; 2000US-0229345P.		
PR	05-SEP-2000; 2000US-0229509P.		
PR	05-SEP-2000; 2000US-0229513P.		
PR	08-SEP-2000; 2000US-0231413P.		
PR	21-SEP-2000; 2000US-0234223P.		
PR	21-SEP-2000; 2000US-0234274P.		
PR	25-SEP-2000; 2000US-0234997P.		
PR	27-SEP-2000; 2000US-0235834P.		
PR	29-SEP-2000; 2000US-0236327P.		
PR	29-SEP-2000; 2000US-0236367P.		
PR	29-SEP-2000; 2000US-0236368P.		
PR	29-SEP-2000; 2000US-0236369P.		
PR	02-OCT-2000; 2000US-0236370P.		
PR	02-OCT-2000; 2000US-0236802P.		
PR	02-OCT-2000; 2000US-0237037P.		
PR	02-OCT-2000; 2000US-0237038P.		
PR	02-OCT-2000; 2000US-0237039P.		
PR	13-OCT-2000; 2000US-0237040P.		
PR	13-OCT-2000; 2000US-0239935P.		
PR	20-OCT-2000; 2000US-0240960P.		
PR	20-OCT-2000; 2000US-0241785P.		
PR	20-OCT-2000; 2000US-0241809P.		
PR	01-NOV-2000; 2000US-0244617P.		
PR	17-NOV-2000; 2000US-0249299P.		
PR	08-DEC-2000; 2000US-0251856P.		
PR	08-DEC-2000; 2000US-0251868P.		
PR	08-DEC-2000; 2000US-0251869P.		
XX			
PA	(ROSE/) ROSEN C A.	RESULT 8	
PA	(RUBE/) RUBEN S M.	AAU16416	
PA	(BARA/) BARASH S C.	ID AAU16416 standard; protein; 292 AA.	
XX		AC AAU16416;	
XX		XX	
PI	Rosen CA, Ruben SM, Barash SC;	XX	
XX		DT 07-NOV-2001 (first entry)	
XX		XX	
DR	WPI; 2003-147444/14.	XX	Human novel secreted protein, Seq ID 1369.
DR	N-P5DB; AEX73306.	XX	
XX		KW	Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
PT	New polypeptides and nucleic acids, useful in gene therapy for treating,	KW	cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
PT	inhibiting or preventing e.g. neural, immune system, muscular,	KW	antibacterial; virucide; fungicide; optalmalogical; vulnery;
PT	respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or	KW	secreted protein; rheumatoid arthritis; hyperproliferative disorder;
PT	renal disorders.	KW	cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX		KW	cerebral ischaemia; angiogenesis; nervous system disorder;
XX		KW	Alzheimer's disease; infection; ocular disorder; corneal infection;
XX		KW	wound healing; epithelial cell proliferation; skin ageing; food additive;
XX		XX	preservative; antiproliferative.



XX PI Rosen CA, Bazash SC, Ruben SM;  
 XX WPI; 2001-488783/53.  
 DR N-PSDB; AAS26403.  
 XX  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 PS Claim 11; SEQ ID NO 1369; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Antibodies to the proteins can also be used in  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi, and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed

Query Match 20.7%; Score 161; DB 4; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-153;  
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 354 VAMQEMECPSIKKRAWETILDKRLPPPTFGSQPTLQTLRWGTNDKSTAPIAKP 413  
 DB 122 VAMQEMECPSIKKRAWETILDKRLPPPTFGSQPTLQTLRWGTNDKSTAPIAKP 181

OY 414 LATRNSESLHGNKPGSVKPTQTAVKESLTDLQTRKEDTPNENKQKRIFFQFLYNN 473  
 DB 182 LATRNSESLHGNKPGSVKPTQTAVKESLTDLQTRKEDTPNENKQKRIFFQFLYNN 241

OY 474 NTRQQTAEADDLHCPWCTLNCRKLYSLKHLKLCNRSRPIFN 514  
 DB 242 NTRQQTAEADDLHCPWCTLNCRKLYSLKHLKLCNRSRPIFN 282

RESULT 9  
 ABUS5485  
 ID ABUS5485 standard; protein; 292 AA.  
 XX ABUS5485;  
 XX  
 XX 18-MAR-2003 (first entry)  
 XX  
 DE Human novel polypeptide #572.  
 XX  
 KW Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cycostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX  
 OS Homo sapiens.

XX US2002132753-A1.  
 XX 19-SEP-2002.  
 XX  
 XX 17-JAN-2001; 2001US-00764864.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-022824P.  
 PR 01-SEP-2000; 2000US-0228287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 06-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 XX Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI; 2003-147444/14.  
 XX N-PSDB; ABX73744.  
 XX  
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 PS Claim 11; SEQ ID NO 1369; 402pp; English.

XX The invention relates to human novel polypeptides and their associated  
CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
CC therapy for treating, inhibiting or preventing neural disorders, immune  
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and  
CC ABU55748 represent human novel polypeptides of the invention  
XX  
SQ Sequence 292 AA;  
Query Match 20.7%; Score 161; DB 6; Length 292;  
Best Local Similarity 100.0%; Pred. No. 1.5e-153;  
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 354 VAMQWESCPISKKATWETILDGKRLPPFPFSGQPTLQTLRWGTNDKSTAPIAKP 413  
Db 122 VAMQWESCPISKKATWETILDGKRLPPFPFSGQPTLQTLRWGTNDKSTAPIAKP 181  
QY 414 LATRNSESLHQENKFGSVKPTQTTIAVKESLTTDLQTRKEKTPNENRKLRFYQFLYNN 473  
Db 182 LATRNSESLHQENKFGSVKPTQTTIAVKESLTTDLQTRKEKTPNENRKLRFYQFLYNN 241  
QY 474 NTRQTEARDLHCPWCTLNCRKLYSLKLKHLKCHSRFIFN 514  
Db 242 NTRQTEARDLHCPWCTLNCRKLYSLKLKHLKCHSRFIFN 282  
RESULT 10  
AAU16402  
ID AAU16402 standard; protein; 175 AA.  
XX  
AC AAU16402;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human novel secreted protein, Seq ID 1355.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
XX preservative; antiproliferative.  
OS Homo sapiens.  
XX  
FN WO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001341.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220663P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225677P.  
PR 14-AUG-2000; 2000US-0225688P.  
PR 14-AUG-2000; 2000US-0225700P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234597P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.

20-OCT-2000; 2000US-0241786P.  
 20-OCT-2000; 2000US-0241787P.  
 20-OCT-2000; 2000US-0241808P.  
 20-OCT-2000; 2000US-0241809P.  
 20-OCT-2000; 2000US-0241826P.  
 01-NOV-2000; 2000US-0244617P.  
 08-NOV-2000; 2000US-0246474P.  
 08-NOV-2000; 2000US-0246475P.  
 08-NOV-2000; 2000US-0246476P.  
 08-NOV-2000; 2000US-0246477P.  
 08-NOV-2000; 2000US-0246478P.  
 08-NOV-2000; 2000US-0246523P.  
 08-NOV-2000; 2000US-0246524P.  
 08-NOV-2000; 2000US-0246525P.  
 08-NOV-2000; 2000US-0246526P.  
 08-NOV-2000; 2000US-0246527P.  
 08-NOV-2000; 2000US-0246528P.  
 08-NOV-2000; 2000US-0246532P.  
 08-NOV-2000; 2000US-0246609P.  
 08-NOV-2000; 2000US-0246610P.  
 08-NOV-2000; 2000US-0246611P.  
 08-NOV-2000; 2000US-0246613P.  
 17-NOV-2000; 2000US-0249208P.  
 17-NOV-2000; 2000US-0249209P.  
 17-NOV-2000; 2000US-0249210P.  
 17-NOV-2000; 2000US-0249211P.  
 17-NOV-2000; 2000US-0249212P.  
 17-NOV-2000; 2000US-0249213P.  
 17-NOV-2000; 2000US-0249214P.  
 17-NOV-2000; 2000US-0249215P.  
 17-NOV-2000; 2000US-0249216P.  
 17-NOV-2000; 2000US-0249217P.  
 17-NOV-2000; 2000US-0249218P.  
 17-NOV-2000; 2000US-0249244P.  
 17-NOV-2000; 2000US-0249245P.  
 17-NOV-2000; 2000US-0249264P.  
 17-NOV-2000; 2000US-0249265P.  
 17-NOV-2000; 2000US-0249297P.  
 17-NOV-2000; 2000US-0249299P.  
 17-NOV-2000; 2000US-0249300P.  
 01-DEC-2000; 2000US-0250160P.  
 01-DEC-2000; 2000US-0250391P.  
 05-DEC-2000; 2000US-0251030P.  
 05-DEC-2000; 2000US-0251988P.  
 05-DEC-2000; 2000US-0256719P.  
 06-DEC-2000; 2000US-0251479P.  
 08-DEC-2000; 2000US-0251856P.  
 08-DEC-2000; 2000US-0251868P.  
 08-DEC-2000; 2000US-0251869P.  
 08-DEC-2000; 2000US-0251989P.  
 08-DEC-2000; 2000US-0251990P.  
 11-DEC-2000; 2000US-0254097P.  
 05-JAN-2001; 2001US-0259678P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-488783/53.  
 N-PSDB; AAS26389.  
 New nucleic acid molecules encoding 461 human secreted proteins for  
 diagnosing, preventing, treating or ameliorating medical conditions and  
 used as food additives or preservatives.  
 Claim 11; SEQ ID NO 1355; 980pp; English.  
 The invention relates to isolated nucleic acid molecules and their  
 encoded secreted proteins. The nucleic acids and proteins are used to  
 prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 in diagnosing a pathological condition or susceptibility to a

CC pathological condition. Antibodies to the proteins can also be used in  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC (ELISA). Disorders which are diagnosed or enzyme linked immunosorbant assays  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 Query Match 16.9%; Score 131; DB 4; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-123;  
 Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 470 LYNNTRQOTEARDDHCPWCTLNCRKLYSLKHLKCHSRFIFNYVHPKGARIDVSN 529  
 DB 5 LYNNTRQOTEARDDHCPWCTLNCRKLYSLKHLKCHSRFIFNYVHPKGARIDVSN 64  
 QY 530 ECVDSYAGNPQDIHQPGAFSRNGPVKRTPTITILVCRPKTKASMSFESEDEGEVE 589  
 DB 65 ECVDSYAGNPQDIHQPGAFSRNGPVKRTPTITILVCRPKTKASMSFESEDEGEVE 124  
 QY 590 QORTYSSGHNH 600  
 DB 125 QORTYSSGHNH 135  
 RESULT 11  
 ABUS5471  
 ID ABUS5471 standard; protein; 175 AA.  
 AC ABUS5471;  
 XX  
 DT 18-MAR-2003 (first entry)  
 XX  
 DE Human novel polypeptide #558.  
 XX  
 KW Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2002132753-A1.  
 XX  
 XX 19-SEP-2002.  
 XX  
 XX 17-JAN-2001; 2001US-00764864.  
 XX  
 XX 31-JAN-2000; 2000US-0179065P.  
 XX 04-FEB-2000; 2000US-0180628P.  
 XX 28-JUN-2000; 2000US-0214886P.  
 XX 07-JUL-2000; 2000US-0216647P.  
 XX 07-JUL-2000; 2000US-0216880P.  
 XX 11-JUL-2000; 2000US-0217487P.  
 XX 14-JUL-2000; 2000US-0217496P.  
 XX 16-JUL-2000; 2000US-0218290P.  
 XX 26-JUL-2000; 2000US-0220963P.  
 XX 26-JUL-2000; 2000US-0220964P.

SQ Sequence 175 AA;

Query Match 16.9%; Score 131; DB 6; Length 175;  
Best Local Similarity 100.0%; Pred.No.1.9e-123;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LYNNNTSQOTEARDLLHCPWCTLNCRLKLYSLKLHLKLCHSRFFINYYVHPKGARIDYSIN 529  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 5 LYNNNTSQOTEARDLLHCPWCTLNCRLKLYSLKLHLKLCHSRFFINYYVHPKGARIDYSIN 64

QY 530 ECYDGSYAGNPQDIHRQPGFAFRNGPVKRTPTIHLVCRPKRTKASMSFLESEEDGEVE 589  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 65 ECYDGSYAGNPQDIHRQPGFAFRNGPVKRTPTIHLVCRPKRTKASMSFLESEEDGEVE 124

QY 590 QORTYSSGHN R 600  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 125 QORTYSSGHN R 135

RESULT 12  
ADB64250  
ID ADB64250 standard; protein; 243 AA.  
XX  
AC ADB64250;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human protein encoded by clone DFNES20076340.  
XX  
KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
KW cell regeneration; membrane protein; signal transduction-related protein;  
KW transcription-related protein; osteoporosis; neurological disease;  
KW cancer; tumour.  
XX  
OS Homo sapiens.  
XX  
PN EP1308459-A2.  
XX  
PD 07-MAY-2003.  
XX  
PF 26-MAR-2002; 2002EP-00007401.  
XX  
PR 05-NOV-2001; 2001JP-00373298.  
PR 25-JAN-2002; 2002US-00350978.  
XX  
(HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Iseino Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX  
WPI: 2003-450961/43.  
DR N-PSDB; ADB62280.  
DR  
XX  
PT New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
XX  
XX  
Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related

CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours). The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a protein of the invention. Note: Some of the  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.

SQ Sequence 243 AA;  
 Query Match 16.5%; Score 128; DB 7; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-120; Indels 0; Gaps 0;  
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60  
 DB 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60  
 QY 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120  
 DB 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120  
 QY 121 SFRSSTPT 128  
 DB 121 SFRSSTPT 128

RESULT 13  
 ADC06802  
 ID ADC06802 standard; protein; 243 AA.  
 XX AC ADC06802;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE Human prostate cancer-related protein PCP0840.  
 XX KW cytostatic; prostate cancer; breast; gene therapy; transgenic; human;  
 XX KW chromosome 7p15.1.  
 XX OS Homo sapiens.  
 XX FN WO2003064599-A2.  
 XX PD 07-AUG-2003.  
 XX PF 24-JAN-2003; 2003WO-US001943.  
 XX PR 25-JAN-2002; 2002US-00054935.  
 XX PR 14-FEB-2002; 2002US-0356130P.  
 XX PR 22-MAR-2002; 2002US-00102946.  
 XX PR 08-APR-2002; 2002US-00117229.  
 XX PR 14-MAY-2002; 2002US-00144198.  
 XX PR 19-JUL-2002; 2002US-00197824.  
 XX PA (ORIG-) ORIGENE TECHNOLOGIES INC.  
 XX PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;  
 XX WPI; 2003-679495/64.  
 XX DR New isolated polynucleotide related to cancer genes, useful for  
 XX PT detecting, diagnosing, staging, monitoring, prognosticating, preventing  
 XX PT or treating cancers, e.g. breast and prostate cancers.  
 XX PS Claim 5; Fig 6; 128pp; English.  
 XX CC The invention relates to a novel isolated polynucleotide comprising a

CC differentially-regulated mammalian cancer gene. The polynucleotides of  
 CC the invention demonstrate cytostatic activity and are differentially  
 CC expressed in prostate cancer. The polynucleotide, polypeptides and  
 CC methods of the invention may be useful for detecting, diagnosing,  
 CC staging, monitoring, prognosticating, preventing or treating cancers,  
 CC particularly breast and prostate cancers. Furthermore, the invention may  
 CC be utilised during gene therapy procedures or in the production of  
 CC transgenic animals. The current sequence is that of the prostate cancer-  
 CC related protein of the invention.

SQ Sequence 243 AA;  
 Query Match 16.5%; Score 128; DB 7; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-120; Indels 0; Gaps 0;  
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60  
 DB 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60  
 QY 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120  
 DB 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120  
 QY 121 SFRSSTPT 128  
 DB 121 SFRSSTPT 128

RESULT 14  
 AAU16381  
 ID AAU16381 standard; protein; 278 AA.  
 XX AC AAU16381;  
 XX DT 07-NOV-2001 (first entry)  
 XX DE Human novel secreted protein, Seq ID 1334.  
 XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
 KW preservative; antiproliferative.  
 XX OS Homo sapiens.  
 XX FN WO200155322-A2.  
 XX PD 02-AUG-2001.  
 XX PF 17-JAN-2001; 2001WO-US001341.  
 XX PR 31-JAN-2000; 2000US-0179065P.  
 XX PR 04-FEB-2000; 2000US-0180628P.  
 XX PR 24-FEB-2000; 2000US-0184664P.  
 XX PR 02-MAR-2000; 2000US-0186350P.  
 XX PR 16-MAR-2000; 2000US-0189874P.  
 XX PR 17-MAR-2000; 2000US-0190076P.  
 XX PR 18-APR-2000; 2000US-0198123P.  
 XX PR 19-MAY-2000; 2000US-0205515P.  
 XX PR 07-JUN-2000; 2000US-0209467P.  
 XX PR 28-JUN-2000; 2000US-0214886P.  
 XX PR 30-JUN-2000; 2000US-021513P.  
 XX PR 07-JUL-2000; 2000US-0218647P.  
 XX PR 07-JUL-2000; 2000US-0216880P.  
 XX PR 11-JUL-2000; 2000US-0217487P.  
 XX PR 11-JUL-2000; 2000US-0217496P.  
 XX PR 14-JUL-2000; 2000US-0218290P.



PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226681P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226686P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226827P.	PR	17-NOV-2000;	2000US-0249208P.
PR	23-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	30-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	01-SEP-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0228287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249214P.
PR	05-SEP-2000;	2000US-0229503P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249216P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249218P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0232081P.	PR	17-NOV-2000;	2000US-0249299P.
PR	12-SEP-2000;	2000US-0231969P.	PR	17-NOV-2000;	2000US-0249300P.
PR	14-SEP-2000;	2000US-0233397P.	PR	01-DEC-2000;	2000US-0250160P.
PR	14-SEP-2000;	2000US-0233398P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0233399P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0233400P.	PR	05-DEC-2000;	2000US-0251388P.
PR	14-SEP-2000;	2000US-0233401P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0233603P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251868P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251869P.
PR	21-SEP-2000;	2000US-0234274P.	PR	08-DEC-2000;	2000US-0251989P.
PR	25-SEP-2000;	2000US-0234997P.	PR	08-DEC-2000;	2000US-0251990P.
PR	25-SEP-2000;	2000US-0234998P.	PR	11-DEC-2000;	2000US-0254097P.
PR	26-SEP-2000;	2000US-0235484P.	PR	05-JAN-2001;	2001US-0259678P.
PR	27-SEP-2000;	2000US-0235834P.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235836P.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327P.	XX	WPI; 2001-488783/53.	
PR	29-SEP-2000;	2000US-0236367P.	XX	N-PSDB; AAS26368.	
PR	29-SEP-2000;	2000US-0236368P.	DR	New nucleic acid molecules encoding 461 human secreted proteins for	
PR	29-SEP-2000;	2000US-0236369P.	DR	diagnosing, preventing, treating or ameliorating medical conditions and	
PR	29-SEP-2000;	2000US-0236369P.	XX	used as food additives or preservatives.	
PR	29-SEP-2000;	2000US-0236370P.	PT	Claim 11; SEQ ID NO 1334; 980pp; English.	
PR	29-SEP-2000;	2000US-0236370P.	XX	The invention relates to isolated nucleic acid molecules and their	
PR	29-SEP-2000;	2000US-0236370P.	XX	encoded secreted proteins. The nucleic acids and proteins are used to	
PR	29-SEP-2000;	2000US-0236370P.	XX	prevent, treat or ameliorate a medical condition in e.g. humans, mice,	
PR	29-SEP-2000;	2000US-0236370P.	XX	rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used	
PR	29-SEP-2000;	2000US-0236370P.	XX	in diagnosing a pathological condition or susceptibility to a	
PR	29-SEP-2000;	2000US-0236370P.	XX	pathological condition. Antibodies to the proteins can also be used in	
PR	29-SEP-2000;	2000US-0236370P.	XX	alleviating symptoms associated with the disorders and in diagnostic	
PR	29-SEP-2000;	2000US-0236370P.	XX	immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays	
PR	29-SEP-2000;	2000US-0236370P.	XX	(ELISA). Disorders which are diagnosed or treated include autoimmune	
PR	29-SEP-2000;	2000US-0236370P.	XX	diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.	
PR	29-SEP-2000;	2000US-0236370P.	XX	neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac	
PR	29-SEP-2000;	2000US-0236370P.	XX	arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,	
PR	08-NOV-2000;	2000US-0246474P.	XX		

CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed

Query Match 14.0%; Score 109; DB 4; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-101; Indels 0; Gaps 0;  
 Matches 109; Conservative 0; Mismatches 0;  
 QY 1 MTGIAAASFFNTCRFGCGGLHFTPLADLIEHIEDNHDTPRVLEKQELQQPTVALSY 60  
 DB 36 MTGIAAASFFNTCRFGCGGLHFTPLADLIEHIEDNHDTPRVLEKQELQQPTVALSY 95  
 QY 61 INRFMTDAARREQESLKKIKQPKLSLTSSSVSRGNVSTPPRHSSGSLT 109  
 DB 96 INRFMTDAARREQESLKKIKQPKLSLTSSSVSRGNVSTPPRHSSGSLT 144

RESULT 15  
 ABUS5450  
 ID ABUS5450 standard; protein; 278 AA.  
 XX  
 AC ABUS5450;  
 XX  
 DT 18-MAR-2003 (first entry)  
 XX  
 DE Human novel polypeptide #537.

XX Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.

XX Homo sapiens.  
 OS  
 XX  
 PN US2002132753-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 11-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234977P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239353P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI; 2003-147444/14.  
 DR N-PSDB; ABX73709.  
 XX  
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 XX Claim 11; SEQ ID NO 1334; 402pp; English.

XX The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
 CC ABUS5748 represent human novel polypeptides of the invention

XX Sequence 278 AA;  
 SQ  
 Query Match 14.0%; Score 109; DB 6; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-101; Indels 0; Gaps 0;  
 Matches 109; Conservative 0; Mismatches 0;  
 QY 1 MTGIAAASFFNTCRFGCGGLHFTPLADLIEHIEDNHDTPRVLEKQELQQPTVALSY 60  
 DB 36 MTGIAAASFFNTCRFGCGGLHFTPLADLIEHIEDNHDTPRVLEKQELQQPTVALSY 95  
 QY 61 INRFMTDAARREQESLKKIKQPKLSLTSSSVSRGNVSTPPRHSSGSLT 109

Db 96 INFRMTDARRRQESLKKIQPKLSLTLSSSVSRGCVSTPPRHSSGSLT 144

Search completed: August 25, 2004, 18:11:48  
Job time : 136 secs

100

100

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:25:28 ; Search time 132 Seconds  
(without alignments)  
1661.037 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 4078  
Sequence: 1 MTGIAAASFNTCRFGGCG.....KALETDSVSGVSKSKKQL 776

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3725	91.3	776	5	ABG30843
2	3413	83.7	803	3	AB300066
3	3237	79.4	739	5	ABG30842
4	2043	50.1	388	4	AAU15378
5	2043	50.1	388	6	ABU55047
6	1497	36.7	289	4	AAU15958
7	1497	36.7	289	6	ABU55027
8	1453	35.6	292	4	AAU16416
9	1453	35.6	292	6	ABU55485
10	1127.5	27.6	955	4	ABB71635
11	919	22.5	175	4	AAU16402
12	919	22.5	175	6	ABU55471
13	661	16.2	243	7	ADCB64250
14	661	16.2	243	7	ADCB64250
15	652	16.0	278	4	AAU16381
16	652	16.0	278	6	ABU55450
17	490	12.0	243	5	ABG30841
18	476	11.7	143	4	AAU15931
19	476	11.7	143	6	ABU55000
20	437.5	10.7	102	3	ABG03182
21	429	10.5	202	3	ABG03182
22	280	6.9	171	4	ABU11690
23	238.5	5.8	604	3	AAV53933
24	230	5.6	498	3	AGS54168
25	227.5	5.6	367	3	AGS54169

26	224	5.5	611	3	AAV53932	AAV53932 A MPC1 pr
27	224	5.5	767	4	ABBS8240	ABBS8240 Drosophil
28	221	5.4	445	3	AAB00060	AAB00060 VRN2 poly
29	217	5.3	440	3	AAB00061	AAB00061 VRN2 poly
30	211	5.2	692	6	ABP70524	ABP70524 Histone d
31	211	5.2	813	3	ABO1674	ABO1674 His2 prot
32	176.5	4.3	1142	4	AAG70713	AAG70713 S cerevis
33	176.5	4.3	1142	6	ABRS3123	ABRS3123 Protein s
34	176.5	4.3	1349	4	AAG70854	AAG70854 C albica
35	161.5	4.0	583	5	ABG93227	ABG93227 C. albica
36	159.5	3.9	339	3	AAG54170	AAG54170 Arabidops
37	156.5	3.8	3696	5	ABP40235	ABP40235 Staphyloc
38	153.5	3.8	1279	4	AAG83047	AAG83047 S. epide
39	153.5	3.8	3692	6	ABU43311	ABU43311 Protein e
40	151	3.7	1881	5	ABP73809	ABP73809 Candida a
41	150	3.7	912	2	AAR36731	AAR36731 Ubiquitin
42	149	3.7	1087	2	AAV19935	AAV19935 B. burgdo
43	149	3.7	1119	2	AAV19934	AAV19934 B. burgdo
44	147.5	3.6	1790	6	ABRS3116	ABRS3116 Protein s
45	147	3.6	907	4	ABUS3073	ABUS3073 Intracell

ALIGNMENTS

RESULT 1  
ABG30843  
ID ABG30843 standard; protein; 776 AA.

XX AC ABG30843;

XX DT 21-OCT-2002 (first entry)

XX DE Human jAZF1/jJAZ1 fusion protein.

XX KW Human; jAZF1, juxtaposed with another zinc finger; jJAZ1, jAZF1/jJAZ1, joined with jAZF1; proliferation; endometrial stroma tumour; immunogen; antigen; antibody; fertility; pregnancy; gene therapy; vaccine; fusion protein.

XX OS Homo sapiens.

XX OS Synthetic.

XX OS Chimeric.

XX FH Key Location/Qualifiers

FT Misc-difference 13..37  
/note= "Encoded by ACCTGCCGATTCGGGGCTGGGGACTCCACTTC-CCACCCCTGGCCGACCTCATCGAGCACATGAGGACCAACCAC. C2H2 zinc finger domain"

FT Region  
/note= "Breakpoint for production of fusion protein jAZF1/jJAZ1"

FT Misc-difference 357  
/note= "Encoded by CAG"

FT Misc-difference 485..508  
/note= "Encoded by CTGCATTGCCCTTGGTGTACTCTGAAGTGGCGCAA-ACCTTATAGTTACTCAAGCATCTTAAGCTCTGCCAT. C2H2 zinc finger domain"

FT Region  
/note= "Bipartite nuclear localisation signal"

FT WO200193805-A2.

FT 13-DEC-2001.

FT 04-JUN-2001; 2001WO-US017936.

FT 02-JUN-2000; 2000US-0209093P.

FT (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.

FT Koontz J, Sklar J;

DR WPI; 2002-575047/61.  
 DR N-PSDB; ABR89163.  
 XX Novel JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides useful as immunogens or  
 PT antigens to raise or test anti-JAZF1, jJAZ1 or JAZF1/jJAZ1 antibodies.  
 XX  
 PS Claim 7; Fig 3; 76pp; English.  
 XX  
 CC The present invention relates to a new JAZF1 (juxtaposed with another  
 CC zinc finger), jJAZ1 (joined with JAZF1) or JAZF1/jJAZ1 polypeptide. The  
 CC methods of the invention can be used to identify a compound which  
 CC controls proliferation of endometrial stroma, by expressing jJAZ1 in the  
 CC presence of the compound, and determining whether the compound affects  
 CC expression of jJAZ1. JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides are useful  
 CC as immunogens or antigens to raise or test anti-JAZF1, jJAZ1 or  
 CC JAZF1/jJAZ1 antibodies. The invention can be used as bait proteins in a  
 CC two hybrid assay or three hybrid assay to identify other proteins which  
 CC bind or interact with JAZF1/jJAZ1-binding proteins. JAZF1, jJAZ1 or  
 CC JAZF1/jJAZ1 molecules are useful for identifying the origin of tumour and  
 CC as tumour marker protein to verify that a stromal tumour is from  
 CC endometrium. The antibody is useful for promoting or decreasing fertility  
 CC or pregnancy, and also for treating endometrial stromal tumours. The  
 CC present amino acid sequence represents the human JAZF1/jJAZ1 fusion  
 CC protein of the invention  
 XX  
 SQ Sequence 776 AA;  
 Query Match 91.3%; Score 3725; DB 5; Length 776;  
 Best Local Similarity 93.6%; Pred. No. 1.4e-311;  
 Matches 726; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 1 MTGIAAASFENCTCFGCGGLHFTPLADLIEHDNHDTPRVLEKQELQOPTYVALSY 60  
 DB 1 MTGIAAASFENXXXXXXXXXXXXXXXXXXXXXIDTPRVLEKQELQOPTYVALSY 60  
 QY 61 INRFMTDAARQESLKKKIQKLSLTSSSVRGVSTPPRHSSGSLTPVTPPTPSS 120  
 DB 61 INRFMTDAARQESLKKKIQKLSLTSSSVRGVSTPPRHSSGSLTPVTPPTPSS 120  
 QY 121 SFRSTPTPTQIVRFLRNLIAPIFLHRTLTYMHSRSTNKRKTKVDMLSKVEK 180  
 DB 121 SFRSTPTPTQIVRFLRNLIAPIFLHRTLTYMHSRSTNKRKTKVDMLSKVEK 180  
 QY 181 MKGEQESHLSAHLQLTFTGFHNDKPSNSEQNSVTLEVLVKVCHKKRDXVSCPI 240  
 DB 181 MKGEQESHLSAHLQLTFTGFHNDKPSNSEQNSVTLEVLVKVCHKKRDXVSCPI 240  
 QY 241 RQVPTGKQVPLIDPLNQTGKGNFSLAVSNFEPNSHMVKSYSLLFRVTRGREFN 300  
 DB 241 RQVPTGKQVPLIDPLNQTGKGNFSLAVSNFEPNSHMVKSYSLLFRVTRGREFN 300  
 QY 301 GNINGETNENIDVNEELPARKRNRDEGKTFVAQMTVFDKQRLQLLDGEYEVAMQEME 360  
 DB 301 GNINGETNENIDVNEELPARKRNRDEGKTFVAQMTVFDKQRLQLLDGEYEVAMQEME 360  
 QY 361 ECPISKKEATWETLDGRLPPPTFTFSGQPTLQTLRWGTGNTDKSTAPIAKPLATNSE 420  
 DB 361 ECPISKKEATWETLDGRLPPPTFTFSGQPTLQTLRWGTGNTDKSTAPIAKPLATNSE 420  
 QY 421 SLHQNKPGSVKPTQTIKAVKESLTDLTQTRKEDTPNENKQKRIFYQFLYNNNTROOTE 480  
 DB 421 SLHQNKPGSVKPTQTIKAVKESLTDLTQTRKEDTPNENKQKRIFYQFLYNNNTROOTE 480  
 QY 481 ARDDILHCPWCTNCRKLYSLKHLKCHSRFIFNYVYHPKGRIDVSNICYDGSYAGNP 540  
 DB 481 ARDDXXXXXXXKXXXXXXXXXXXXXXXSRFIFNYVYHPKGRIDVSNICYDGSYAGNP 540  
 QY 541 QDIHQPCGAFSRNGPVKRTPTIHLVCRPKRTKASMSFELEDGEVEQORTYSSGHNH 600  
 DB 541 QDIHQPCGAFSRNGPVKRTPTIHLVCRPKRTKASMSFELEDGEVEQORTYSSGHNH 600  
 QY 601 LYFHSDDTCLPLRPQMEVDSDEKDPWLRKTTIQIEEFSVDNVEGEKVMKLNHLVHK 660

DB 601 LYFHSDDTCLPLRPQMEVDSDEKDPWLRKTTIQIEEFSVDNVEGEKVMKLNHLVHK 660  
 QY 661 HGFIADNQNHACMLFVENYQKIIKQNLCRNFMHLVSMHDFNLISMSIDKAVTKLRE 720  
 DB 661 HGFIADNQNHACMLFVENYQKIIKQNLCRNFMHLVSMHDFNLISMSIDKAVTKLRE 720  
 QY 721 MQOKLEKGSASAPANEITEEQNGTANGFSEINSEKALETDSVSGVSKQSKQKL 776  
 DB 721 MQOKLEKGSASAPANEITEEQNGTANGFSEINSEKALETDSVSGVSKQSKQKL 776  
 RESULT 2  
 AAB00066  
 ID AAB00066 standard; protein; 803 AA.  
 XX AAB00066;  
 XX 16-NOV-2000 (first entry)  
 DT KIAA0160 polypeptide.  
 DE  
 XX Vernalization gene; VRN2; plant characteristic; flowering time;  
 KW leaf size; leaf shape; shade avoidance response; reproduction; breeding;  
 KW pollination; cultivation; human.  
 XX Homo sapiens.  
 XX WO200044918-A1.  
 XX 03-AUG-2000.  
 XX 28-JAN-2000; 2000WO-GB000248.  
 XX 28-JAN-1999; 99GB-00001927.  
 XX (PLAN-) PLANT BIOSCIENCE LTD.  
 XX Dean C, Gendall A;  
 XX WPI; 2000-499333/44.  
 DR N-PSDB; AAA47759.  
 XX Isolated vernalization gene VRN2 is used to produce transgenic plants  
 PT with altered vernalization response, flowering time, leaf size and/or  
 PT shape or shade avoidance response for maximized reproductive success.  
 XX Disclosure; Page 79; 105pp; English.  
 XX Isolated nucleic acid sequences obtained from the VRN2 locus of a plant  
 CC encode polypeptides which are capable of affecting one or more  
 CC vernalization responses such as, flowering time, leaf size and/or shape  
 CC or the shade avoidance response of a plant into which the nucleic acid is  
 CC introduced. Introducing such sequences into plants to alter these  
 CC characteristics maximises the reproductive success of the plant. This  
 CC protein is encoded an isolated human sequence which has homology to the  
 CC Arabidopsis thaliana VRN2 gene over a short region near the N-terminus  
 XX  
 SQ Sequence 803 AA;  
 Query Match 83.7%; Score 3413; DB 3; Length 803;  
 Best Local Similarity 99.8%; Pred. No. 1.2e-284;  
 Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 129 EPTQYRFLRNLIAPIFLHRTLTYMHSRSTNKRKTKVDMLSKVEMKGEQESH 188  
 DB 129 EPTQYRFLRNLIAPIFLHRTLTYMHSRSTNKRKTKVDMLSKVEMKGEQESH 188  
 QY 189 SLSAHLQLTFTGFFHNDKPSNSEQNSVTLEVLVKVCHKKRDXVSCPIRQVPTGKK 248  
 DB 189 SLSAHLQLTFTGFFHNDKPSNSEQNSVTLEVLVKVCHKKRDXVSCPIRQVPTGKK 248  
 QY 216 SUSAHLQLTFTGFFHNDKPSNSEQNSVTLEVLVKVCHKKRDXVSCPIRQVPTGKK 275  
 DB 216 SUSAHLQLTFTGFFHNDKPSNSEQNSVTLEVLVKVCHKKRDXVSCPIRQVPTGKK 275  
 QY 249 QVPLIPDLNQTGKGNFSLAVSNFEPNSHMVKSYSLLFRVTRGREFNMGNETN 308

Db 276 QVPLIPDLNQTGKGNFSPSLAVSSNEFEPSSHMVKSYSLLFRVTRPGRRREFNGMINGETN 335  
 Qy 309 ENIDVNEELPARRKRNEDGKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKRR 368  
 Db 336 ENIDVNEELPARRKRNEDGKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKRR 395  
 Qy 369 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKP 428  
 Db 396 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKP 455  
 Qy 429 GSVKPTQTIAVKESLTDLQTRKEKDTPNENRQKLRIFYQFLYNNTRQOOTEARDDLHCP 488  
 Db 456 GSVKPTQTIAVKESLTDLQTRKEKDTPNENRQKLRIFYQFLYNNTRQOOTEARDDLHCP 515  
 Qy 489 WCTLNCRKLYSLXHLKLSRPFYNYVHPKGAIDVSIINECYDGSVAGNPQDIHQPG 548  
 Db 516 WCTLNCRKLYSLXHLKLSRPFYNYVHPKGAIDVSIINECYDGSVAGNPQDIHQPG 575  
 Qy 549 FAFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVQQRYSGGHNRLYFHSDTC 608  
 Db 576 FAFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVQQRYSGGHNRLYFHSDTC 635  
 Qy 609 LPLRPQEMEVDSEDEKDPWLREKTTIQIEEFSDVNEGEKVMKLNHLVHMKGFIADNQ 668  
 Db 636 LPLRPQEMEVDSEDEKDPWLREKTTIQIEEFSDVNEGEKVMKLNHLVHMKGFIADNQ 695  
 Qy 669 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSDKAVTKLREMOQKLEK 728  
 Db 696 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSDKAVTKLREMOQKLEK 755  
 Qy 729 ESASPANBEITBEOQTANGTANGSEINSKEKALETDSVSGVSKQSKQKL 776  
 Db 756 ESASPANBEITBEOQTANGTANGSEINSKEKALETDSVSGVSKQSKQKL 803

RESULT 3

ID ABG30842 standard; protein; 739 AA.

XX AC ABG30842;

XX DT 21-OCT-2002 (first entry)

XX DE Human joined with jAZF1 (jjAZ1) protein.

XX Human; jAZF1; juxtaposed with another zinc finger; jAZF1; jAZF1/jjAZ1;  
 KW joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;  
 KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;  
 KW chromosome 17.

OS Homo sapiens.

XX Key Location/Qualifiers

XX Region 92..93

XX /note= "Breakpoint for production of fusion protein

XX jAZF1/jjAZ1"

XX Misc-difference 448..471

XX /note= "Encoded by CNGCATGCCCTGTGTACTGTGAACCTGCCGAA-

XX ACTTATAGTTTACTCAAGCATCTAACTCTGCAT. C2H2 zinc finger

XX domain"

XX Region 521..538

XX /note= "Bipartite nuclear localisation signal"

XX WO200193805-A2.

XX 13-DEC-2001.

XX 04-JUN-2001; 2001WO-US017936.

XX 02-JUN-2000; 2000US-0209093P.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.

XX PA

PI Koontz J, Sklar J;  
 XX WPI; 2002-575047/61.  
 DR N-PSDB; ABK89162.  
 XX Novel jAZF1, jjAZ1 or jAZF1/jjAZ1 polypeptides useful as immunogens or  
 PT antigens to raise or test anti-jAZF1, jjAZ1 or jAZF1/jjAZ1 antibodies.  
 XX Claim 5; Fig 2; 76pp; English.  
 XX The present invention relates to a new jAZF1 (juxtaposed with another  
 CC zinc finger), jAZ1 (joined with jAZF1) or jAZF1/jjAZ1 polypeptide. The  
 CC methods of the invention can be used to identify a compound which  
 CC controls proliferation of endometrial stroma, by expressing jjAZ1 in the  
 CC presence of the compound, and determining whether the compound affects  
 CC expression of jAZ, jAZF1, jjAZ1 or jAZF1/jjAZ1 polypeptides are useful  
 CC as immunogens or antigens to raise or test anti-jAZF1, jjAZ1 or  
 CC jAZF1/jjAZ1 antibodies. The invention can be used as bait proteins in a  
 CC two hybrid assay or three hybrid assay to identify other proteins which  
 CC bind or interact with jAZF1/jjAZ1-binding proteins. jAZF1, jjAZ1 or  
 CC jAZF1/jjAZ1 molecules are useful for identifying the origin of tumour and  
 CC as tumour marker protein to verify that a stromal tumour is from  
 CC endometrium. The antibody is useful for promoting or decreasing fertility  
 CC or pregnancy, and also for treating endometrial stromal tumours. The  
 CC present amino acid sequence represents the human jAZ1 protein of the  
 CC invention. This sequence is encoded by the human jAZ1 gene located on  
 CC chromosome 17  
 XX SQ Sequence 739 AA;

Query Match 79.4%; Score 3237; DB 5; Length 739;  
 Best Local Similarity 96.1%; Pred. No. 1.5e-269;  
 Matches 623; Conservative 1; Mismatches 24; Indels 0; Gaps 0;

Qy 129 EPTQIVRFLRTRNLIAPIFLHRTLTYMSHRNSTNKRKTFKVDMLSKVKMKQGQESH 188  
 Db 92 KPTQIVRFLRTRNLIAPIFLHRTLTYMSHRNSTNKRKTFKVDMLSKVKMKQGQESH 151  
 Qy 189 SLSAHLQLTFTGTFHKNKDPSPNSNEQNSVTLEVLVVKVCHKKKDVSCPIRQVPTGKK 248  
 Db 152 SLSAHLQLTFTGTFHKNKDPSPNSNEQNSVTLEVLVVKVCHKKKDVSCPIRQVPTGKK 211  
 Qy 249 QVPLIPDLNQTGKGNFSPSLAVSSNEFEPSSHMVKSYSLLFRVTRPGRRREFNGMINGETN 308  
 Db 212 QVPLIPDLNQTGKGNFSPSLAVSSNEFEPSSHMVKSYSLLFRVTRPGRRREFNGMINGETN 271  
 Qy 309 ENIDVNEELPARRKRNEDGKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKRR 368  
 Db 272 ENIDVNEELPARRKRNEDGKTFVAQMTVFDKNRRLQLLDGEYEVAMQEMEECPISKRR 331  
 Qy 369 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKP 428  
 Db 332 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKP 391  
 Qy 429 GSVKPTQTIAVKESLTDLQTRKEKDTPNENRQKLRIFYQFLYNNTRQOOTEARDDLHCP 488  
 Db 392 GSVKPTQTIAVKESLTDLQTRKEKDTPNENRQKLRIFYQFLYNNTRQOOTEARDDLHCP 451  
 Qy 489 WCTLNCRKLYSLXHLKLSRPFYNYVHPKGAIDVSIINECYDGSVAGNPQDIHQPG 548  
 Db 452 XXXXXXXXXXXXXXXXXXXXSRPFYNYVHPKGAIDVSIINECYDGSVAGNPQDIHQPG 511  
 Qy 549 FAFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVQQRYSGGHNRLYFHSDTC 608  
 Db 512 FAFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVQQRYSGGHNRLYFHSDTC 571  
 Qy 609 LPLRPQEMEVDSEDEKDPWLREKTTIQIEEFSDVNEGEKVMKLNHLVHMKGFIADNQ 668  
 Db 572 LPLRPQEMEVDSEDEKDPWLREKTTIQIEEFSDVNEGEKVMKLNHLVHMKGFIADNQ 631  
 Qy 669 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSDKAVTKLREMOQKLEK 728  
 Db 632 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSDKAVTKLREMOQKLEK 691



QY 729 ESASPANBEITEQNGTANGSEINSEKSKALETDSVSGVSKQKQKL 776  
 Db 692 ESASPANBEITEQNGTANGSEINSEKSKALETDSVSGVSKQKQKL 739

RESULT 4  
 AAU15978  
 ID AAU15978 standard; protein; 388 AA.  
 AC AAU15978;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human novel secreted protein, Seq ID 931.  
 XX  
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 KW cardiac; vasotropic; cerebroprotective; nootropic; neuroprotective;  
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angioneurosis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
 KW preservative; antiproliferative.  
 OS Homo sapiens.  
 XX  
 PN WO20015322-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001341.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225575P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 23-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 08-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234957P.  
 PR 25-SEP-2000; 2000US-0234958P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235835P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.

PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-488783/53.  
N-PSDB; AAS25965.  
New nucleic acid molecules encoding 461 human secreted proteins for  
diagnosing, preventing, treating or ameliorating medical conditions and  
used as food additives or preservatives.  
Claim 11; SEQ ID NO 931; 980pp; English.  
The invention relates to isolated nucleic acid molecules and their  
encoded secreted proteins. The nucleic acids and proteins are used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
in diagnosing a pathological condition or susceptibility to a  
pathological condition. Antibodies to the proteins can also be used in  
alleviating symptoms associated with the disorders and in diagnostic  
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
(ELISA). Disorders which are diagnosed or treated include autoimmune  
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
nervous system disorders e.g. Alzheimer's disease, infections caused by  
bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
and many other disorders listed in the specification. The polypeptides  
can also be used to aid wound healing and epithelial cell proliferation,  
to prevent skin aging due to sunburn, to maintain organs before  
transplantation, for supporting cell culture of primary tissues, to  
regenerate tissues and in chemotaxis. The polypeptides can also be used  
as a food additive or preservative to increase or decrease storage  
capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
minerals, cofactors and other nutritional components. The present  
sequence represents a novel secreted protein of the invention. Note: The  
sequence data for this patent did not form part of the printed  
Query Match 50.1%; Score 2043; DB 4; Length 388;  
Best Local Similarity 99.5%; Pred. No. 4.9e-167;  
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
389 GPTLOFTLWGTGNDKSTAPAKPLATNSESLSHQENKPGSVKPTQIAVKESLTIDLO 448  
1 GPTLOFTLWGTGNDKSTAPAKPLATNSESLSHQENKPGSVKPTQIAVKESLTIDLO 60  
449 TRKEKDTNPNRQKLRIFYQFLYNNNTROQTEARDLHCPWCTINCRKLYSLKHLKLCH 508

Db 61 TRKEKDTNPNRQKLRIFYQFLYNNNTROQTEARDLHCPWCTINCRKLYSLKHLKLCH 120  
Qy 509 SRFIFNYVHPKRGARIDVSINECVGDSYAGNPQDIHRQPGFAFSRNGPVRKRTPIITHLVC 568  
Db 121 SRFIFNYVHPKRGARIDVSINEGVDSYAGNPQDIHRQPGFAFSRNGPVRKRTPIITHLVC 180  
Qy 569 RPKRTKASMEFLESEDEGEVEQORTYSSGHNRILYFHSIDTCLPLRPQMEVDSDEKDPW 628  
Db 181 RPKRTKASMEFLESEDEGEVEQORTYSSGHNRILYFHSIDTCLPLRPQMEVDSDEKDPW 240  
Qy 629 LREKTIQIEBFSDVNEGEKEVMKLNHLVHKHGFADNOMNHACMLFVENVYQKIIKKN 688  
Db 241 LREKTIQIEBFSDVNEGEKEVMKLNHLVHKHGFADNOMNHACMLFVENVYQKIIKKN 300  
Qy 689 LCRNFMHLVSMHDFNLISIMSDKAVTKLREHQKLEKESASPANEETEEQNGTANG 748  
Db 301 LCRNFMHLVSMHDFNLISIMSDKAVTKLREHQKLEKESASPANEETEEQNGTANG 360  
Qy 749 FSEINSKEKALETDSVSGVSKQSKQKL 776  
Db 361 FSEINSKEKALETDSVSGVSKQSKQKL 388  
RESULT 5  
ABUS5047  
ID ABUS5047 standard; protein; 388 AA.  
XX AC ABUS5047;  
XX DT 18-MAR-2003 (first entry)  
XX DE Human novel polypeptide #134.  
KW Human; neural disorder; immune system disorder; renal disorder;  
KW muscular disorder; respiratory disease; reproductive disorder;  
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
KW haemostatic; antiarteriosclerotic.  
OS Homo sapiens.  
XX US2002132753-A1.  
XX 19-SEP-2002.  
XX 17-JAN-2001; 2001US-00764864.  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 16-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239355P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
PA (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX  
PI Rosen CA, Ruben SM, Barash SC;  
XX  
XX WPI; 2003-147444/14.  
DR N-PSDB; ABX73306.  
XX  
XX New polypeptides and nucleic acids, useful in gene therapy for treating,  
PT inhibiting or preventing e.g. neural, immune system, muscular,  
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
PT renal disorders.  
XX  
PS Claim 11; SEQ ID NO 931; 402bp; English.  
XX  
XX The invention relates to human novel polypeptides and their associated  
CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
CC therapy for treating, inhibiting or preventing neural disorders, immune  
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and  
CC ABU55748 represent human novel polypeptides of the invention  
XX  
SQ Sequence 388 AA;  
Query Match 50.1%; Score 2043; DB 6; Length 388;  
Best Local Similarity 99.5%; Pred. No. 4.9e-167;  
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 389 GPTLQFTLRWGETNDKSTAPIAFLATRNSESLHQNKGKSVKPTQTIAVKESLTLDLQ 448  
Db 1 GPTLQFTLRWGETNDKSTAPIAFLATRNSESLHQNKGKSVKPTQTIAVKESLTLDLQ 60  
QY 449 TRKEKDTNENRQKLRIFYQLYNNNTNQTEARDLHCPWCTLNCRLKLYSLNKLKCH 508  
Db 61 TRKEKDTNENRQKLRIFYQLYNNNTNQTEARDLHCPWCTLNCRLKLYSLNKLKCH 120

QY 509 SRFIENYVHPKGAIDVINSINECYDGSYAGNPQDIHQPGFAPSRNGPVKRTPTTHILVC 568  
Db 121 SRFIENYVHPKGAIDVINSINEGYDGSYAGNPQDIHQPGFAPSRNGPVKRTPTTHILVC 180  
QY 569 RPKRTKASMSFELESDGEVEQORTYSSGHNLRYFHSDTCLPLRPOEMEVSEDEKDPW 628  
Db 181 RPKRTKASMSFELESDGEVEQORTYSSGHNLRYFHSDTCLPLRPOEMEVSEDEKDPW 240  
QY 629 LREKTIITQIEEFSVDNKEGEKVMKLNHLVHVKHGFADNQMNHACMLFVNYGQKIKN 688  
Db 241 LREKTIITQIEEFSVDNKEGEKVMKLNHLVHVKHGFADNQMNHACMLFVNYGQKIKN 300  
QY 689 LCRNFMHLVSMHDFNLISIMSIDKAVTKLRMOQKLEKESASPANEEITEEQNGTANG 748  
Db 301 LCRNFMHLVSMHDFNLISIMSIDKAVTKLRMOQKLEKESASPANEEITEEQNGTANG 360  
QY 749 FSEINSKEKALSTDSVSGVSKQSKQKL 776  
Db 361 FSEINSKEKALSTDSVSGVSKQSKQKL 388  
RESULT 6  
AAU15958  
ID AAU15958 standard; protein; 289 AA.  
XX  
AC AAU15958;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human novel secreted protein, Seq ID 911.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
WO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001341.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0186874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0223268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 16-AUG-2000; 2000US-0226273P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0233968P.  
 PR 14-SEP-2000; 2000US-0233977P.  
 PR 14-SEP-2000; 2000US-0233988P.  
 PR 14-SEP-2000; 2000US-0233998P.  
 PR 14-SEP-2000; 2000US-0234000P.  
 PR 14-SEP-2000; 2000US-0234001P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 13-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.

PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249399P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-488783/53.  
 N-PSDB; AAS25945.

New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives.

Claim 11; SEQ ID NO 911; 980pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage

CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 Query Match 36.7%; Score 1497; DB 4; Length 289;  
 Best Local Similarity 99.6%; Pred. No. 3.7e-120;  
 Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 493 NCRKLYSLKHLKLSHSPFIENYVHPKGRIDVINECDGVSAGNPQDIHQPGFAPS 552  
 DB 6 DCRKLYSLKHLKLSHSPFIENYVHPKGRIDVINECDGVSAGNPQDIHQPGFAPS 65  
 QY 553 RNPVKTPTIHLVCRPKRTKASMEFLESGEVEQQRTYSSGHNLVYFSDTCLPLR 612  
 DB 66 RNPVKTPTIHLVCRPKRTKASMEFLESGEVEQQRTYSSGHNLVYFSDTCLPLR 125  
 QY 613 PQMEVDSDEKDPKLEKTIITOLEEFSVNEGEKVMKLNHLVYKRGFIADNQNHA 672  
 DB 126 PQMEVDSDEKDPKLEKTIITOLEEFSVNEGEKVMKLNHLVYKRGFIADNQNHA 185  
 QY 673 CMLFVNYGQKIKNLCRNFMLHVSMDNLSIMSIDKAVTKLREMOQKLEKGSAS 732  
 DB 186 CMLFVNYGQKIKNLCRNFMLHVSMDNLSIMSIDKAVTKLREMOQKLEKGSAS 245  
 QY 733 PANBEITEQNTANGFSEINSKEKALETDSVSGVSKQSKKL 776  
 DB 246 PANBEITEQNTANGFSEINSKEKALETDSVSGVSKQSKKL 289  
 RESULT 7  
 ABUS5027  
 ID ABUS5027 standard; protein; 289 AA.  
 XX AC ABUS5027;  
 XX DT 18-MAR-2003 (first entry)  
 XX DE Human novel polypeptide #114.  
 XX KW Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX OS Homo sapiens.  
 XX PN US2002132753-A1.  
 XX PD 19-SEP-2002.  
 XX PF 17-JAN-2001; 2001US-00764864.  
 XX PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX Rosen CA, Ruben SM, Barash SC;  
 XX WPI; 2003-147444/14.  
 DR N-PSDB; ABX73286.  
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX Claim 11; SEQ ID NO 911; 402pp; English.  
 PS The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma) blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
 CC ABUS5748 represent human novel polypeptides of the invention  
 XX Sequence 289 AA;  
 SQ  
 Query Match 36.7%; Score 1497; DB 6; Length 289;  
 Best Local Similarity 99.6%; Pred. No. 3.7e-120;  
 Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 493 NCRKLYSLKHLKLSHSPFIENYVHPKGRIDVINECDGVSAGNPQDIHQPGFAPS 552

Db :||||| 6 DCKLYSLKHLKCLHSRIFNYVHPKCARIDVSIINECYDGSYAGNPQDIHRQGFQAFS 65  
QY 553 RNPVVKRTITHLVCRPKRTKASMEPLESEDEGEVEQORTYSSGHNRLYFHSDTCLPLR 612  
Db 66 RNPVVKRTITHLVCRPKRTKASMEPLESEDEGEVEQORTYSSGHNRLYFHSDTCLPLR 125  
QY 613 POEMEVDSEKDPLEWREKTTIQIBEFSDVNEGEKVMKLNLMHVMKHGFTADNQMHHA 672  
Db 126 POEMEVDSEKDPLEWREKTTIQIBEFSDVNEGEKVMKLNLMHVMKHGFTADNQMHHA 185  
QY 673 CMLFVENYQKILKNCLEFNLHVLVSMHDFNLISIMSIDKAVTKLREMOQKLEKGESAS 732  
Db 186 CMLFVENYQKILKNCLEFNLHVLVSMHDFNLISIMSIDKAVTKLREMOQKLEKGESAS 245  
QY 733 PANEBITEPQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
Db 246 PANEBITEPQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 289  
  
RESULT 8  
AAU16416  
ID AAU16416 standard; protein; 292 AA.  
XX  
AC AAU16416;  
XX  
DT 07-NOV-2001 (first entry)  
DE Human novel secreted protein, Seq ID 1369.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiac; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
FN WO20015322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001341.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190078P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216890P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234232P.  
PR 21-SEP-2000; 2000US-0234233P.  
PR 25-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234597P.  
PR 26-SEP-2000; 2000US-0234998P.  
PR 27-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239355P.  
PR 13-OCT-2000; 2000US-0239357P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-024617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.

08-NOV-2000; 2000US-0246528P.  
08-NOV-2000; 2000US-0246532P.  
08-NOV-2000; 2000US-0246609P.  
08-NOV-2000; 2000US-0246610P.  
08-NOV-2000; 2000US-0246611P.  
08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249212P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249242P.  
17-NOV-2000; 2000US-0249245P.  
17-NOV-2000; 2000US-0249264P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
01-DEC-2000; 2000US-0250391P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
06-DEC-2000; 2000US-0256719P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-02559678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-488783/53.  
N-P5DB; AAS26403.  
New nucleic acid molecules encoding 461 human secreted proteins for  
diagnosing, preventing, treating or ameliorating medical conditions and  
used as food additives or preservatives.  
Claim 11; SEQ ID NO 1369; 980pp; English.  
The invention relates to isolated nucleic acid molecules and their  
encoded secreted proteins. The nucleic acids and proteins are used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
in diagnosing a pathological condition or susceptibility to a  
pathological condition. Antibodies to the proteins can also be used in  
alleviating symptoms associated with the disorders and in diagnostic  
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
(ELISA). Disorders which are diagnosed or treated include autoimmune  
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,  
nervous system disorders e.g. Alzheimer's disease, infections caused by  
bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
and many other disorders listed in the specification. The polypeptides  
can also be used to aid wound healing and epithelial cell proliferation,  
to prevent skin aging due to sunburn, to maintain organs before  
transplantation, for supporting cell culture of primary tissues, to  
regenerate tissues and in chemotaxis. The polypeptides can also be used  
as a food additive or preservative to increase or decrease storage  
capabilities, fat content, lipid, protein, carbohydrate, vitamins,

CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention. Note: The  
CC sequence data for this patent did not form part of the printed

Query Match 35.6%; Score 1453; DB 4; Length 292;  
Best Local Similarity 96.8%; Pred. No. 2.3e-116;  
Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 233 KDVSCPIRQVPTGKKQVPLPDLNQTGPNFPPSLAVSSNEFSPNSHMVKSYSLLFRVT 292  
Db 1 KDVSCPIRQVPTGKKQVPLPDLNQTGPNFPPSLAVSSNEFSPNSHMVKSYSLLFRVT 60

QY 293 RGRREFNGMNGETNENIDVNEELPARRKKNREDGKTFVAQMTVFQKNERLQLLDGEY 352

Db 61 RGRREFNGMNGETNENIDVNEELPARRKKNREDGKTFVAQMTVFQKNERLQLLDGEY 120

QY 353 EVAMQEMEECPISKRAWTETILDGKRLPPFETFSQGTLOFTLRWGTNDKSTAPIAK 412

Db 121 XVAMQEMEECPISKRAWTETILDGKRLPPFETFSQGTLOFTLRWGTNDKSTAPIAK 180

QY 413 PLATNSSELHONKPGSVKPTQTIAVKESLTTDLQTRKEKDTNENRQKLRIFVQFLYN 472

Db 181 PLATNSSELHONKPGSVKPTQTIAVKESLTTDLQTRKEKDTNENRQKLRIFVQFLYN 240

QY 473 NNTRQOTEARDDLHCPWCTLNCRKLYSLKHLKCHSRFIFNYV 516

Db 241 NNTRQOTEARDDLHCPWCTLNCRKLYSLKHLKCHSRFIFNYV 284

## RESULT 9

ABU55485

ID ABU55485 standard; protein; 292 AA.

XX AC ABU55485;

XX DT 18-MAR-2003 (first entry)

XX DE Human novel polypeptide #572.

XX KW Human; neutral disorder; immune system disorder; renal disorder;

XX KW muscular disorder; respiratory disease; reproductive disorder;

XX KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;

XX KW hyperproliferative disorder; inflammatory disease; allergic reaction;

XX KW blood related disorder; cancer; immunosuppressive; antineoplastic;

XX KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;

XX KW haemostatic; antiarteriosclerotic.

XX OS Homo sapiens.

XX PN US2002132753-A1.

XX PD 19-SEP-2002.

XX PF 17-JAN-2001; 2001US-00764864.

XX PF 31-JAN-2000; 2000US-0179065P.

XX PF 04-FEB-2000; 2000US-0180628P.

XX PF 28-JUN-2000; 2000US-0214886P.

XX PF 07-JUL-2000; 2000US-0216647P.

XX PF 07-JUL-2000; 2000US-0216880P.

XX PF 11-JUL-2000; 2000US-0217487P.

XX PF 11-JUL-2000; 2000US-0217496P.

XX PF 14-JUL-2000; 2000US-0218290P.

XX PF 26-JUL-2000; 2000US-0220963P.

XX PF 14-AUG-2000; 2000US-0224518P.

XX PF 14-AUG-2000; 2000US-0224519P.

XX PF 14-AUG-2000; 2000US-0225267P.

XX PF 14-AUG-2000; 2000US-0225268P.

XX PF 14-AUG-2000; 2000US-0225270P.

XX PF 14-AUG-2000; 2000US-0225447P.

XX PF 14-AUG-2000; 2000US-0225757P.

XX PF 14-AUG-2000; 2000US-0225758P.



PR 22-AUG-2000; 2000US-0228968P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI; 2003-147444/14.  
 DR N-PSDB; ABX73744.  
 XX  
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular, or  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 PS Claim 11; SEQ ID NO 1369; 402pp; English.  
 XX  
 CC The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory reactions and conditions (e.g. asthma), blood  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and  
 CC ABU55748 represent human novel polypeptides of the invention  
 XX  
 SQ Sequence 292 AA;  
 Query Match 35.6%; Score 1453; DB 6; Length 292;  
 Best Local Similarity 96.8%; Pred. No. 2.3e-116;  
 Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 QY 233 KDVSCPIQVPTGKQVPLPDLNQTGKFPSPSLAVSSNFEPSXSHMVKSYLLPRVT 292  
 :|||||

Db 1 KDVSCPIQVPTGKQVPLPDLNQTGKFPSPSLAVSSNFEPSXSHMVKSYLLPRVT 60  
 QY 293 RQGRREFNGMNGTNIENIDVNEELPARRKXNREDEGKTFVAQMTVDFKXRLQLDGEY 352  
 Db 61 RQGRREFNGMNGTNIENIDVNEELPARRKXNREDEGKTFVAQMTVDFKXRLQLDGEY 120  
 QY 353 EVAMQEMEBCPISKKRATWETILDGKRLPPFETFSQGTTLQFTLRWTGETNDKSTAPIAK 412  
 Db 121 XVAQEMEBCPISKKRATWETILDGKRLPPFETFSQGTTLQFTLRWTGETNDKSTAPIAK 180  
 QY 413 PLATRNSES LHQENKPGSVKPTQTIAVKESLTTDLQTRKEKOTPNENRQKLRFYQPLYN 472  
 Db 181 PLATRNSES LHQENKPGSVKPTQTIAVKESLTTDLQTRKEKOTPNENRQKLRFYQPLYN 240  
 QY 473 NNTROQTEARDDLHCPWCTLNCRLKLSLLKHLKLSRRTFNVV 516  
 Db 241 NNTROQTEARDDLHCPWCTLNCRLKLSLLKHLKLSRRTFNVV 284  
 RESULT 10  
 ABB71635  
 ID ABB71635 standard; protein; 955 AA.  
 XX  
 AC ABB71635;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 41697.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL15738.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 41697; 2lpp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 955 AA;  
 Query Match 27.6%; Score 1127.5; DB 4; Length 955;  
 Best Local Similarity 32.9%; Pred. No. 1.8e-87;  
 Matches 273; Conservative 137; Mismatches 240; Indels 179; Gaps 26;

QY 65 MTDARREQESLKKIKQKLSLTLSSVSRRGNV-----STPRHSSGSLTPVTPPTPSS 120  
Db 1 MAPAKREKDS-----NPDGSAANGILGTHGAPDASNAGSTVPTAGQVKLN 49  
QY 121 SFRSSTP-----TEPTQIYRFLTRNLIAPIFLHRTLYTMYSHRNRNINIKRTKFKYDDM 174  
Db 50 GHQOQELFLQAFKPTQIYLRNRHETNPIFLNRTLSYMKERMSRNNKKRISFQVNSM 109  
QY 175 LSKVEMKGEQESHLSA-HLQFTGFFHKNDKPSSENEQ-----NSVTL 222  
Db 110 LESI-----TOKSEAVSONYLHVYDSLHEKLPARLDNESGEDLLQQLCEAGESVSE 164  
QY 223 VLLVVKCHKRDKVSCPTQVPTGKQVPLIPDLNQTGPNFPLAVSSNEFEP-SNSHM 281  
Db 165 TLYKITSKKDSTLDQELLKSCQIVYNP---KORVGEHATISIPLOTMRPMGEQHT 221  
QY 282 VKYSLLFRVTRPGRRENGINGETNENIDVNEELPARRKRNREDGEKTKVAQMTVDFK 341  
Db 222 L--YKLLERIK-----VLSPSTCN--DENAETPP-NKRSRPN-EKMGSELIYK 266  
QY 342 NERLQDGEYEVAMQEECPI---SKRATWETILDGKLP---PREFSQGPTI-LFT 395  
Db 267 SSGP-ITGEVEAMQLPNSISIKSFSPKCTWTMPD-SYPLSLDYDYVQSPMLKFH 324  
QY 396 LRWTGETNDKSTAPIAKPLATR-----NSLSHOEN-----KPGSVKPTQTI 437  
Db 325 LTL---SNEQLPEMISAPELQRYVQHLDAVAMNYYNNNNNNCSGLKNGSGGNSV 381  
QY 438 AVKESLTTDQTRKEKDPNENRQKLRIFYOLYNNNTROTEARDLHCPWCITLNCRL 497  
Db 382 C-----KTPP-----EHQIVYFMYSNNTRQTEYTQELNCPWGLDCLRL 423  
QY 498 YSLKHLKLCHSRFIFNTVYHPKARIDVSNIECYDGSYAGNPQDIHRQPGFAPSRN-GP 556  
Db 424 YALLKHLKLCHARNFTYQAGSGARIDVTINDAYDGSYAGSPYDLAGPSSSPARTCGP 483  
QY 557 VKRTPITHLVCRPRTKASMEFLESDGEVQOPTYSSGHNRLYFHSOTCLPLRQEM 616  
Db 484 VRRTSVTSLMCRPRQKTCLEFLEDEDEISNQRSYITGHNRLYHHTETCLPVHPKEL 543  
QY 617 EVDSEKDPWLREKTTIQTIEPSDVNKEGVKMLNHLVHKHGFIDNMNHAQWLF 676  
Db 544 DIDSEGSDDLWLRQKTTQMDIDESDVNKEGKELMLNHLVHMGHGFVGCQLPIACEMP 603  
QY 677 VYENYQKIKKNCNFMHLVSHVDHFNLSIMSDIKAVTKLRMQOKLEKGES----- 730  
Db 604 LDKAGTEIVRKNLYENFILHCSLFDYGLIAETVYKTVQKLOGLLSKYAAGBELMQOR 663  
QY 731 -----ASPANEITE----- 740  
Db 664 BEQLKWLVDGMHKQEDPKTKSPQKPPADQASTSSASTSGSGSGSSSQMPKRMPPA 723  
QY 741 -----EONGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
Db 724 HLKRGSAASSPGVQSGKTENGTSNS-----SSNSKQVAKKSADQPL 767

## RESULT 11

AAU16402

ID AAU16402 standard; protein; 175 AA.

XX AC AAU16402;

XX DT 07-NOV-2001 (first entry)

DE DE Human novel secreted protein, Seq ID 1355.

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiac; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;

KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
XX preservative; antiproliferative.

OS Homo sapiens.

XX WO200155322-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US0001341.

XX PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 06-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 08-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.

PR 12-SEP-2000; 2000US-0232081P.

PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.

PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0241826P.  
 PR 08-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0244674P.  
 PR 08-NOV-2000; 2000US-024475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 17-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250319P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 08-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-488783/53.  
 XX N-PSDB; AAS26389.  
 DR New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX Claim 11; SEQ ID NO 1355; 980pp; English.  
 XX The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Antibodies to the proteins can also be used in  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 Query Match 22.5%; Score 919; DB 4; Length 175;  
 Best Local Similarity 94.3%; Pred. No. 1.1e-70;  
 Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
 QY 466 FYQFLYNNTRQQTTEARDLLHCPWCTLNCRKLYSLKHLKLCRSRTFNVVHPKGARID 525  
 DB 1 FYQFLYNNTRQQTTEARDLLHCPWCTLNCRKLYSLKHLKLCRSRTFNVVHPKGARID 60  
 QY 526 VSINECYDGSYAGNPQDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESD 585  
 DB 61 VSINECYDGSYAGNPQDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESD 120  
 QY 586 GEVEQORTYSSGHNRXLYFHSDDTCLPQPQMEVDSDEKDPWLREKTIQTIEEP 640  
 DB 121 GEVEQORTYSSGHNRXLYFHSDDTCLPQPQMEVDSDEKDPWLREKTIQTIEEP 175  
 RESULT 12  
 ABUS5471  
 ID ABUS5471 standard; protein; 175 AA.  
 XX AC ABUS5471;  
 XX DT 18-MAR-2003 (first entry)  
 XX Human novel polypeptide #558.  
 XX Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;

KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX  
 XX Homo sapiens.

XX US2002132753-A1.

XX 19-SEP-2002.

XX 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 28-JUN-2000; 2000US-0214886P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 22-AUG-2000; 2000US-0225758P.

XX 20-AUG-2000; 2000US-0226868P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 01-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229509P.

XX 05-SEP-2000; 2000US-0229513P.

XX 08-SEP-2000; 2000US-0231413P.

XX 21-SEP-2000; 2000US-0234223P.

XX 21-SEP-2000; 2000US-0234274P.

XX 21-SEP-2000; 2000US-0234997P.

XX 27-SEP-2000; 2000US-0235834P.

XX 29-SEP-2000; 2000US-0236327P.

XX 29-SEP-2000; 2000US-0236367P.

XX 29-SEP-2000; 2000US-0236368P.

PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 XX Claim 11; SEQ ID NO 1355; 402pp; English.

XX The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
 CC ABUS5748 represent human novel polypeptides of the invention

XX Sequence 175 AA;

Query Match 22.5%; Score 919; DB 6; Length 175;

Best Local Similarity 94.3%; Pred. No. 1.1e-70;

Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 466 FYQFLYNNNTROQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIENYVHPKGARID 525

Db 1 FYQFLYNNNTROQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIENYVHPKGARID 60

QY 526 VSINECYDGSYAGNPQDIHQPGFAFSRNGPVKRTPIITHILVCRPKTKASMSFELED 585

Db 61 VSINECYDGSYAGNPQDIHQPGFAFSRNGPVKRTPIITHILVCRPKTKASMSFELED 120

QY 586 GEVEQRTYSSGHNRILYFHSDFCLPLRQMEVDSDEKDPWLRKTIQIEEF 640

Db 121 GEVEQRTYSSGHNRILYFHSDFCLPLRQMEVDSDEKDPWLRKTIQIEEF 175

RESULT 13

ADB64250

XX ID ADB64250 standard; protein; 243 AA.

XX ADB64250;

XX DT 04-DEC-2003 (first entry)

XX Human protein encoded by clone DFNES20076340.

XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumour.

XX Homo sapiens.

XX EP1308459-A2.

XX 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.

XX 05-NOV-2001; 2001JP-00379298.

XX 25-JAN-2002; 2002US-00350978.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuho Y;

XX New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,

DR WPI; 2003-450961/43.  
XX N-PSDB; ADB62280.  
XX  
PT New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
XX  
XX  
PS Claim 1; Page; 222pp; English.  
XX  
XX The invention discloses a polynucleotide comprising a sequence selected  
CC from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or  
CC peptide of the polynucleotide by contacting the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesising the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotides and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumours). The CDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a protein of the invention. Note: Some of the  
CC sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.  
XX  
SQ Sequence 243 AA;

Query Match 16.2%; Score 661; DB 7; Length 243;  
Best Local Similarity 100.0%; Pred. No. 3.4e-48;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHDNHDTPRVLEKQELQOPTYVALSY 60  
DB 1 MTGIAAASFNTCRFGGGLHPTLADLIEHDNHDTPRVLEKQELQOPTYVALSY 60  
QY 61 INRFMTDAARREQESLKKIQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVPPTPSS 120  
DB 61 INRFMTDAARREQESLKKIQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVPPTPSS 120  
QY 121 SFRSSTPT 128  
DB 121 SFRSSTPT 128

RESULT 14  
ADC06802  
ID ADC06802 standard; protein; 243 AA.  
XX  
AC ADC06802;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human prostate cancer-related protein PCP0840.  
XX  
KW cytostatic; prostate cancer; breast; gene therapy; transgenic; human;  
KW chromosome 7p15.1.  
XX  
OS Homo sapiens.  
XX  
PN WO2003064599-A2.  
XX  
PD 07-AUG-2003.  
XX  
PF 24-JAN-2003; 2003WO-US001943.

XX 25-JAN-2002; 2002US-00054935.  
PR 14-FEB-2002; 2002US-0356130P.  
PR 22-MAR-2002; 2002US-00102946.  
PR 08-APR-2002; 2002US-00117229.  
PR 14-MAY-2002; 2002US-00144198.  
PR 19-JUL-2002; 2002US-00197824.  
XX  
PA (ORIG-) ORIGENE TECHNOLOGIES INC.  
XX  
XX Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;  
XX WPI; 2003-679495/64.  
XX  
XX New isolated polynucleotide related to cancer genes, useful for  
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing  
PT or treating cancers, e.g. breast and prostate cancers.  
XX  
XX Claim 5; Fig 6; 128pp; English.  
PS  
XX The invention relates to a novel isolated polynucleotide comprising a  
CC differentially-regulated mammalian cancer gene. The polynucleotides of  
CC the invention demonstrate cytostatic activity and are differentially  
CC expressed in prostate cancer. The polynucleotide, polypeptides and  
CC methods of the invention may be useful for detecting, diagnosing,  
CC staging, monitoring, prognosticating, preventing or treating cancers,  
CC particularly breast and prostate cancers. Furthermore, the invention may  
CC be utilised during gene therapy procedures or in the production of  
CC transgenic animals. The current sequence is that of the prostate cancer-  
CC related protein of the invention.  
XX  
SQ Sequence 243 AA;

Query Match 16.2%; Score 661; DB 7; Length 243;  
Best Local Similarity 100.0%; Pred. No. 3.4e-48;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHDNHDTPRVLEKQELQOPTYVALSY 60  
DB 1 MTGIAAASFNTCRFGGGLHPTLADLIEHDNHDTPRVLEKQELQOPTYVALSY 60  
QY 61 INRFMTDAARREQESLKKIQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVPPTPSS 120  
DB 61 INRFMTDAARREQESLKKIQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVPPTPSS 120  
QY 121 SFRSSTPT 128  
DB 121 SFRSSTPT 128

RESULT 15  
AAU16381  
ID AAU16381 standard; protein; 278 AA.  
XX  
AC AAU16381;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human novel secreted protein, Seq ID 1334.  
XX  
DE Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulvaray;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
XX WO200155322-A2.  
PN

XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001341.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-MAR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226273P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228927P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUYA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX

DR WPI; 2001-488783/53.  
DR N-PSDB; AAS26368.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
FT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives.  
XX  
XX Claim 11; SEQ ID NO 1334; 980pp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Antibodies to the proteins can also be used in  
CC alleviating symptoms associated with the disorders and in diagnostic  
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays  
CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
CC and many other disorders listed in the specification. The polypeptides  
CC can also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
Query Match 16.0%; Score 652; DB 4; Length 278;  
Best Local Similarity 99.2%; Pred No. 2.5e-47;  
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MTGIAAAGFFSNTCRFGGGLHFTPLADLIEHDNHDIDTPRVLEKQELQPTYVALSY 60  
Db |||||  
36 MTGIAAAGFFSNTCRFGGGLHFTPLADLIEHDNHDIDTPRVLEKQELQPTYVALSY 95  
QY 61 INRFMTDAARREQESLKKIOPKLSLTLSVVSGNVSTPPRHSSGSLTPVTPPTPSS 120  
Db |||||  
96 INRFMTDAARREQESLKKIOPKLSLTLSVVSGNVSTPPRHSSGSLTPVTPPTPSS 155  
QY 121 SFRSSTPT 128  
Db |||||  
156 SFRSSTPT 163

Search completed: August 25, 2004, 17:54:35  
Job time : 137 secs





GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:06:05 ; Search time 134 Seconds  
(without alignments)  
1827.180 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 776  
Sequence: 1 MTGIAAASFNTCRFGCG.....KALETDSVSGVSKQKQKL 776

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_arched.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	1170	5 Q95T15	Q95t15 drosophila
2	9	1.2	1990	5 Q9U8Q0	Q9u8q0 drosophila
3	9	1.2	1995	5 Q9W244	Q9w244 drosophila
4	8	1.0	75	10 Q9STQ5	Q9stq5 arabidopsis
5	8	1.0	93	5 Q61719	Q61719 anopheles g
6	8	1.0	150	10 Q3FP19	Q3fp19 oryza sativ
7	8	1.0	150	10 Q8H2H9	Q8h2h9 oryza sativ
8	8	1.0	156	10 Q9LEY9	Q9ley9 arabidopsis
9	8	1.0	214	10 P33540	P33540 silene prat
10	8	1.0	214	10 Q9ST39	Q9st39 silene lati
11	8	1.0	217	10 O04301	O04301 silene lati
12	8	1.0	217	10 P33689	P33689 silene lati
13	8	1.0	247	12 Q8B9D8	Q8b9d8 rachiplusia
14	8	1.0	281	4 Q9EQW4	Q9eqw4 homo sapien
15	8	1.0	281	16 Q8XBK6	Q8xbk6 escherichia
16	8	1.0	297	16 Q8NPG7	Q8np7 corynebacte

17	8	1.0	305	16 Q81E61	Q81e61 bacillus ce
18	8	1.0	307	17 Q8U0U5	Q8u0u5 pyrococcus
19	8	1.0	316	10 Q9LS48	Q9ls48 arabidopsis
20	8	1.0	340	4 Q9BRF4	Q9brf4 homo sapien
21	8	1.0	342	4 Q9Y6J7	Q9y6j7 homo sapien
22	8	1.0	385	4 Q9BQW7	Q9bqw7 homo sapien
23	8	1.0	442	17 Q9UYR0	Q9uyr0 pyrococcus
24	8	1.0	447	16 Q895X3	Q895x3 clostridium
25	8	1.0	476	13 Q13256	Q13256 gallus gall
26	8	1.0	478	5 Q95YB9	Q95yb9 caenorhabdi
27	8	1.0	503	16 Q9CLB3	Q9clb3 pasteurellia
28	8	1.0	506	13 Q9DEF6	Q9def6 gallus gall
29	8	1.0	508	13 Q9DEF5	Q9def5 gallus gall
30	8	1.0	514	13 Q13255	Q13255 gallus gall
31	8	1.0	520	4 Q81XL3	Q81xl3 homo sapien
32	8	1.0	525	10 Q8H133	Q8h133 arabidopsis
33	8	1.0	526	10 Q98ZQ3	Q98zq3 arabidopsis
34	8	1.0	668	4 Q9BR66	Q9br66 homo sapien
35	8	1.0	697	16 Q7VQM0	Q7vqm0 candidatus
36	8	1.0	881	5 Q81SG6	Q81sg6 stylonychia
37	8	1.0	3525	5 Q81SW5	Q81sw5 plasmodium
38	7	0.9	51	16 Q8FPP5	Q8fpp5 corynebacte
39	7	0.9	51	16 Q7ULM1	Q7ulm1 rhodospirell
40	7	0.9	52	16 P74799	P74799 synechocyst
41	7	0.9	56	12 Q77N22	Q77n22 human picob
42	7	0.9	70	16 Q8DD29	Q8dd29 vibrio vuln
43	7	0.9	76	16 Q81HA1	Q81ha1 bacillus ce
44	7	0.9	84	16 Q87HG7	Q87hg7 vibrio para
45	7	0.9	88	16 Q8R7X6	Q8r7x6 thermoaer

# ALIGNMENTS

## RESULT 1

Q95T15	PRELIMINARY;	PRT;	1170 AA.
ID Q95T15	PRELIMINARY;	PRT;	1170 AA.
AC Q95T15	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT 01-DEC-2001	(TrEMBLrel. 19, Last sequence update)		
DT 01-OCT-2002	(TrEMBLrel. 22, Last annotation update)		
DB SD01586P.			
GN PX OR CG4444.			
OS Drosophila melanogaster (Fruit fly).			
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC Ephydroidea; Drosophilidae; Drosophila.			
OX NCBI_TaxID=7227;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Scapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,			
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,			
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,			
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.			
EL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AY058753; AAL13982.1; -			
DR FlyBase; FBgn0003175; px.			
SQ SEQUENCE 1170 AA; 124080 MW; 37F9331CF6891DD3 CRC64;			

Query Match 1.2%; Score 9; DB 5; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	113	TPPTPSSS 121
Db	907	TPPTPSSS 915

## RESULT 2

Q9U8Q0	PRELIMINARY;	PRT;	1990 AA.
ID Q9U8Q0	PRELIMINARY;	PRT;	1990 AA.
AC Q9U8Q0			

DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE PLEXUS.  
 GN PX OR CG4444.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=20025921; PubMed=10556047;  
 RA Matkatsuo H., Tadokoro R., Gamo S., Hayashi S.;  
 RT "Regression of the wing vein development in Drosophila by the nuclear  
 RT matrix protein Plexus.";  
 RL Development 126:5207-5216(1999).  
 DR EMBL; AB032181; BAA86717.1; -;  
 DR FlyBase; FBgn0003175; px.  
 SQ SEQUENCE 1990 AA; 213113 MW; 5707C4F2986DA9E2 CRC64;  
  
 Query Match 1.2%; Score 9; DB 5; Length 1990;  
 Best Local Similarity 100.0%; Pred.No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 113 TPPITPSSS 121  
 Db 1727 TPPITPSSS 1735  
  
 RESULT 3  
 Q9W244  
 ID Q9W244 PRELIMINARY; PRT; 1995 AA.  
 AC Q9W244;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE CG4444 protein.  
 GN PX OR CG4444.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Beran B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaramila M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RP FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AS003457; AAF46855.2; -;  
 DR FlyBase; FBgn0003175; px.  
 SQ SEQUENCE 1995 AA; 213862 MW; E7A563275E1D8C7 CRC64;  
  
 Query Match 1.2%; Score 9; DB 5; Length 1995;  
 Best Local Similarity 100.0%; Pred.No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 113 TPPITPSSS 121  
 Db 1732 TPPITPSSS 1740  
  
 RESULT 4  
 Q9STQ5  
 ID Q9STQ5 PRELIMINARY; PRT; 75 AA.  
 AC Q9STQ5;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-WAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN T25K17.100 OR AT4G26290.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Koetter P., Hempel S., Entian K.-D., Bancroft I.,  
 RA Mewes H.W., Mayer K.F.X., Schueller C.;  
 RN Submitted (MAR-1999) to the EMBL/GenBank/DBSJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RN Submitted (MAR-1999) to the EMBL/GenBank/DBSJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
 RL [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
 RL EMBL; AL049171; CAB38958.1; -  
 DR EMBL; AL161564; CAB79484.1; -  
 DR PIR; T06013; T06013.  
 KW Hypothetical protein.  
 SQ SEQUENCE 75 AA; 8649 MW; 584259469EBD2BD1 CRC64;  
 Query Match 1.0%; Score 8; DB 10; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 8.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 115 PITPSSSF 122  
 DB 48 PITPSSSF 55  
 RESULT 5  
 O61719 PRELIMINARY; PRT; 93 AA.  
 AC O61719;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.  
 OX NCBI\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=G3;  
 RA Romans P., Law A.K.-T., Binns K., Han Y.-S., Cornel A.J.,  
 RA Vasiliauskas D., Paskewitz S.M.;  
 RT "Molecular genetics of catecholamine metabolism in the malaria vector  
 RT Anopheles gambiae: analysis of the dopa decarboxylase gene."  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AF063021; AAC16248.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 93 AA; 10569 MW; 8E1BBA930E41AD84 CRC64;  
 Query Match 1.0%; Score 8; DB 5; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 LSLTSSSS 91  
 DB 46 LSLTSSSS 53  
 RESULT 6  
 Q9FP19 PRELIMINARY; PRT; 150 AA.  
 ID Q9FP19  
 AC Q9FP19;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE P0038C05.4 protein.  
 GN P0038C05.4.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC  
 RT clone:P0038C05.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AP003044; BAB19331.1; -  
 DR HSSP; P55769; 127K.  
 DR Gramene; Q9FP19; -  
 DR GO; GO:0005840; C:ribosome; IEA.  
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.  
 DR GO; GO:0006412; P:protein biosynthesis; IEA.  
 DR InterPro; IPR004038; Ribosomal\_L7A.  
 DR InterPro; IPR004037; Ribosomal\_L7Ae.  
 DR Pfam; PF01248; Ribosomal\_L7Ae; 1.  
 DR PRINTS; PRO0881; L7ARS6FAMILY.  
 SQ SEQUENCE 150 AA; 16354 MW; 6F626F759AE30108 CRC64;  
 Query Match 1.0%; Score 8; DB 10; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 408 APIAKPLA 415  
 DB 17 APIAKPLA 24  
 RESULT 7  
 Q8H2H9 PRELIMINARY; PRT; 150 AA.  
 AC Q8H2H9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE P0676F10.14 protein.  
 GN P0676F10.14.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC  
 RT clone:P0676F10.";  
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AP005813; BAC20622.1; -  
 DR GO; GO:0005840; C:ribosome; IEA.  
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.  
 DR GO; GO:0006412; P:protein biosynthesis; IEA.  
 DR InterPro; IPR004038; Ribosomal\_L7A.  
 DR InterPro; IPR004037; Ribosomal\_L7Ae.  
 DR Pfam; PF01248; Ribosomal\_L7Ae; 1.  
 DR PRINTS; PRO0881; L7ARS6FAMILY.  
 SQ SEQUENCE 150 AA; 16354 MW; 6F626F759AE30108 CRC64;  
 Query Match 1.0%; Score 8; DB 10; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 408 APIAKPLA 415

```

Db      17 APIAKPLA 24
|||||||
RESULT 8
Q9LEY9 PRELIMINARY; PRT; 156 AA.
AC Q9LEY9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Nhp2-like protein (AF59g08180/T22D6_120).
GN T22D6.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL "Arabidopsis cDNA clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Chung M.K., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yu G., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL "Arabidopsis cDNA clones.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP EMBL; AL357612; CAB93719.1; -
DR EMBL; AY050477; AAK91490.1; -
DR EMBL; AF378885; AAK55688.1; -
DR FIR; T50503; T50503.
DR HSP; P55769; 1E7K.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; P:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR004038; Ribosomal_L7A.
DR InterPro; IPR004037; Ribosomal_L7Ae.
DR Pfam; PF0248; Ribosomal_L7Ae; 1.
DR PRINTS; PR00891; L7ARSG6FAMILY.
SQ SEQUENCE 156 AA; 16948 MW; CC8B7C196E92442A CRC64;

Query Match 1.0%; Score 8; DB 10; Length 156;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 APIAKPLA 415
Db 23 APIAKPLA 30
|||||||
RESULT 9
P93540 PRELIMINARY; PRT; 214 AA.
AC P93540;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MROS3 protein precursor.
GN MROS3.
OS Silene pratensis (White campion) (Lychnis alba).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=52853;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Anther;
RX MEDLINE=97048713; PubMed=8893544;
RA Matsunaga S., Kawano S., Takano H., Uchida H., Sakai A., Kuroiwa T.;
RT "Isolation and developmental expression of male reproductive organ-
RT specific genes in a dioecious campion, Melandrium album (Silene
RT latifolia).";
RL Plant J. 10:679-689(1996).
DR EMBL; D82027; BAA11513.1; -
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
RW Signal.
FT SIGNAL.
FT SIGNAL.
SQ SEQUENCE 214 AA; 23461 MW; 1931BC0C91935AFB CRC64;

Query Match 1.0%; Score 8; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSILTSSS 91
Db 16 LSILTSSS 23
|||||||
RESULT 10
Q9ST39 PRELIMINARY; PRT; 214 AA.
AC Q9ST39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MROS3A.
GN MROS3A.
OS Silene latifolia (White campion).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX Matsunaga S., Kawano S., Kuroiwa T.;
RT "Chromosomal localization of MROS3 genomic clones in a dioecious
RT campion, Silene latifolia.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=97048713; PubMed=8893544;
RA Matsunaga S., Kawano S., Takano H., Uchida H., Sakai A., Kuroiwa T.;
RT "Isolation and developmental expression of male reproductive organ-
RT specific genes in a dioecious campion, Melandrium album (Silene
RT latifolia).";
RL Plant J. 10:679-689(1996).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=SATI; TISSUE=Leaf;
RX MEDLINE=20040083; PubMed=10571898;
RA Matsunaga S., Schutze K., Donnison I.S., Grant S.R., Kuroiwa T.,
RA Kawano S.;
RT "Single pollen typing combined with laser-mediated manipulation.";
RL Plant J. 20:371-378(1999).
DR EMBL; AB013611; BAA84010.1; -.
DR EMBL; AB029398; BAA90868.1; -.
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
SQ SEQUENCE 214 AA; 23461 MW; 492EAF0480CD7C20 CRC64;

Query Match 1.0%; Score 8; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTSSSS 91
Db 16 LSLTSSSS 23

RESULT 11
O04301
ID O04301 PRELIMINARY; PRT; 217 AA.
AC O04301;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Men-9a protein.
GN MEN-9A.
OS Silene latifolia (white campion).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=flower.
RA Robertson S.E., Scutt C.P., Li Y., Willis M.E., Gilmartin P.M.;
RT "Expression dynamics of the men-9 genes delimit the third whorl in
RT dioecious Silene latifolia flowers.";
RL Plant J. 0:0-0(0).
DR EMBL; Y12324; CAA72993.1; -.
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
SQ SEQUENCE 217 AA; 23946 MW; 2E3640753B266F0E CRC64;

Query Match 1.0%; Score 8; DB 10; Length 217;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTSSSS 91
Db 17 LSLTSSSS 24

RESULT 12
P93689
ID P93689 PRELIMINARY; PRT; 217 AA.
AC P93689;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CCL54 protein.
GN CCL54.
OS Silene latifolia (white campion).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Anther;

RX MEDLINE=98042964; PubMed=9375394;
RA Barbacar N., Himmelsdaels S., Farbos I., Moneger F., Lardon A.,
RA Delichere C., Mouras A., Negrutiu I.;
RT "Isolation of early genes expressed in reproductive organs of the
RT dioecious white campion (Silene latifolia) by subtraction cloning
RT using an asexual mutant";
RL Plant J. 12:805-817(1997).
DR EMBL; X94358; CAA64143.1; -.
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
SQ SEQUENCE 217 AA; 23968 MW; 154F8B059BB80D84 CRC64;

Query Match 1.0%; Score 8; DB 10; Length 217;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTSSSS 91
Db 17 LSLTSSSS 24

RESULT 13
Q8B9D8
ID Q8B9D8 PRELIMINARY; PRT; 247 AA.
AC Q8B9D8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Rachiplusia ou multiple nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=80366;
RN [1]
RP SEQUENCE FROM N.A.
RA Bonning B.C., Harrison R.L.;
RT "The Rachiplusia ou multiple nucleopolyhedrovirus genome sequence.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY145471; AAN28079.1; -.
KW Hypothetical protein.
SQ SEQUENCE 247 AA; 28705 MW; FD6943F62F4D7A9 CRC64;

Query Match 1.0%; Score 8; DB 12; Length 247;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 YVALSYIN 62
Db 23 YVALSYIN 30

RESULT 14
Q9BQW4
ID Q9BQW4 PRELIMINARY; PRT; 281 AA.
AC Q9BQW4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DJ811H13.1.4 (Phospholipase C, beta 4 (1-phosphatidylinositol-4,5-
DE bisphosphate phosphodiesterase beta 4) (Isoform 4)) (Fragment).
GN PLCB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Blakey S.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL023805; CAC34529.1; -.
DR HSSP; P10688; 1DJX.
DR GO; GO:0004435; P:phosphoinositide phospholipase C activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.

```

```

DR GO: 0006629; P: lipid metabolism; IEA.
DR InterPro: IPR001192; PI-PLC.
DR InterPro: IPR000909; PI-PLC_Xdom.
DR Pfam: PF00388; PI-PLC-X_1.
DR PRINTS: PR00390; PHPLIPASEC.
DR SMART: SM00148; PLCXC; 1.
DR PROSITE: PS00007; PIP2C_X_DOMAIN; 1.
FT NON TER 281
SQ SEQUENCE 281 AA; 32062 MW; 66549A020E7D7F6F CRC64;

Query Match 1.0%; Score 8; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 GESASpan 735
DB 216 GESASpan 223

RESULT 15
Q8XB6 PRELIMINARY; PRT; 281 AA.
AC Q8XB6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative transposase (St46 protein).
GN ST46 OR 24330 OR ECS3859.
OS Escherichia coli, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=83/39;
RA Tauschek M., Strugnell R.A., Robins-Browne R.M.;
RT "Characterization of the LEE pathogenicity islands of rabbit
enteropathogenic Escherichia coli.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=413/89-1;
RA Benkel P., Chakraborty T.;
RT "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
producing Escherichia coli.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RP SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
Apodaca J., Arantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[4]
SEQUENCE FROM N.A.
RP SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL: AF453441; AAL57563.1; -
DR EMBL: AJ277443; CAC81894.1; -
DR EMBL: AE005528; AAG58114.1; -

```

```

DR EMBL: AP002563; BAB37282.1; ALT_INIT.
DR PIR: C91111; C91111.
DR PIR: P85956; P85956.
DR GO: 0003677; F: DNA binding; IEA.
DR GO: 0006310; P: DNA recombination; IEA.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; rve; 1.
KW Complete proteome.
SQ SEQUENCE 281 AA; 32894 MW; 1A64154BB8067514 CRC64;

Query Match 1.0%; Score 8; DB 16; Length 281;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SLKHLKL 506
DB 223 SLKHLKL 230

```

Search completed: August 25, 2004, 18:14:44  
Job time : 141 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:07:10 ; Search time 46 Seconds  
(without alignments)  
1622.709 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 776  
Sequence: 1 MTGTAASFFSNTCRFGCG.....KALETDSVSGVSKQKQL 776

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:\*  
1: Pir1:\*  
2: Pir2:\*  
3: Pir3:\*  
4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.0	75	T05013	hypothetical prote
2	8	1.0	156	T50503	nbp2-like protein
3	8	1.0	247	E72865	AcOrf-124 protein
4	8	1.0	262	C91111	hypothetical prote
5	8	1.0	281	F85956	probable transposa
6	8	1.0	442	C75057	hypothetical prote
7	8	1.0	526	T03541	hypothetical prote
8	8	1.0	641	JC7331	gamma-glutamyltran
9	8	1.0	669	T08827	hypothetical prote
10	7	0.9	46	1 VTVA2	viscotoxin A2 - Eu
11	7	0.9	46	1 VFDP1T	phoratoxin - Calif
12	7	0.9	52	S75642	hypothetical prote
13	7	0.9	105	A31046	SusE protein limpo
14	7	0.9	109	F91132	hypothetical prote
15	7	0.9	116	H90626	NADH dehydrogenase
16	7	0.9	129	T0282	hypothetical prote
17	7	0.9	142	E84040	hypothetical prote
18	7	0.9	144	T50649	elicitor-responsiv
19	7	0.9	152	1 QPCE1	regulatory protein
20	7	0.9	152	F91214	transcription regu
21	7	0.9	152	F8060	regulator for asna
22	7	0.9	152	2 AC0953	regulatory protein
23	7	0.9	153	1 Q4ADB2	early E4 17K prote
24	7	0.9	154	2 T50100	Nucleolar protein,
25	7	0.9	154	2 T43644	nbp2 homolog - fis
26	7	0.9	155	2 B70030	conserved hypothet
27	7	0.9	157	2 S75905	hypothetical prote
28	7	0.9	164	2 T01726	hypothetical prote
29	7	0.9	190	2 AB2305	6,7-dimethyl-8-rib

30	7	0.9	198	2 B81389	hypothetical prote
31	7	0.9	201	2 T46404	hypothetical prote
32	7	0.9	222	2 C75539	conserved hypothet
33	7	0.9	223	2 T23356	hypothetical prote
34	7	0.9	224	2 D96643	hypothetical prote
35	7	0.9	227	2 AH2123	hypothetical prote
36	7	0.9	230	2 B86824	two-component syst
37	7	0.9	234	2 S22134	capsid protein - r
38	7	0.9	237	2 S75120	transposase sll198
39	7	0.9	239	2 T45806	hypothetical prote
40	7	0.9	247	2 F84420	hypothetical prote
41	7	0.9	251	2 E96637	hypothetical prote
42	7	0.9	254	2 T15952	hypothetical prote
43	7	0.9	260	2 T36955	hypothetical prote
44	7	0.9	269	2 T35647	phosphomethylpyrim
45	7	0.9	269	2 AG0718	probable toxin-lik

ALIGNMENTS

RESULT 1  
T06013  
hypothetical protein T25K17.100 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
C:Accession: T06013  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15382  
A:Accession: T06013  
A:Molecule type: DNA  
A:Residues: 1-75 <BEV>  
A:Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.100  
A:Experimental source: cultivar Columbia; BAC clone T25K17  
C:Genetics:  
A:Gene: ATSP:T25K17.100  
A:Map position: 4

Query Match 1.0%; Score 8; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 PTPSSSF 122  
| | | | | | | |  
Db 48 PTPSSSF 55

RESULT 2  
T50503  
nbp2-like protein - Arabidopsis thaliana  
N:Alternate names: protein T22D6.120  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: T50503  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25101  
A:Accession: T50503  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-156 <BEV>  
A:Cross-references: EMBL:AL357612  
A:Experimental source: cultivar Columbia; BAC clone T22D6  
C:Genetics:  
A:Map position: 5  
A:Introns: 45/1; 68/2; 103/3  
A:Note: T22D6.120

Query Match 1.0%; Score 8; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 APIAKPLA 415  
 |||||  
 Db 23 APIAKPLA 30

## RESULT 3

AcOrf-124 protein - Autographa californica nuclear polyhedrosis virus  
 C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV  
 A:Note: dsDNA virus  
 C:Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 12-Nov-1999  
 C:Accession: E72865  
 R:AYres, W.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.  
 Virology 202, 566-605, 1994  
 A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.  
 A:Reference number: A72850; MUID:94303173; PMID:8030224  
 A:Accession: E72865  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-247 <AYR>  
 A:Cross-references: GB:L22858; NID:gs10708; PIDN:AAA66754.1; PID:gs559193  
 C:Genetics:  
 A:Gene: ACOrf-124

Query Match 1.0%; Score 8; DB 2; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 9.4; 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0;

QY 55 YVALSYIN 62  
 |||||  
 Db 23 YVALSYIN 30

## RESULT 4

hypothetical protein ECS3859 [imported] - Escherichia coli (strain O157:H7, substrain R1  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: C9111  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
 A:Reference number: A39629; MUID:21156231; PMID:11258796  
 A:Accession: C9111  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-262 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BAE37282.1; PID:gl3363331; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain R1MD 0509952  
 C:Genetics:  
 A:Gene: ECS3859

Query Match 1.0%; Score 8; DB 2; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 9.9; 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0;

QY 499 SLLKHLKL 506  
 |||||  
 Db 204 SLLKHLKL 211

## RESULT 5

Probable transposase 24330 [imported] - Escherichia coli (strain O157:H7, substrain ED18  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: F85956  
 R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: F85956  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-281 <STO>  
 A:Cross-references: GB:AE005174; NID:gl2517538; PIDN:AAG58114.1; GSPDB:GN00145; UWGP:Z433;  
 A:Experimental source: strain O157:H7, substrain ED1933  
 C:Genetics:  
 A:Gene: Z4330

Query Match 1.0%; Score 8; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SLLKHLKL 506  
 |||||  
 Db 223 SLLKHLKL 230

## RESULT 6

hypothetical protein PAB1420 - Pyrococcus abyssi (strain Orsay)  
 C:Species: Pyrococcus abyssi  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: C75057

R:anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
 A:Reference number: A75001  
 A:Accession: C75057  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-442 <KAW>  
 A:Cross-references: GB:AJ248287; GB:AL096836; NID:gs458657; PIDN:CAB50352.1; PID:el516255  
 A:Experimental source: strain Orsay  
 C:Genetics:  
 A:Gene: PAB1420

Query Match 1.0%; Score 8; DB 2; Length 442;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 LLDGEYEV 354  
 |||||  
 Db 135 LLDGEYEV 142

## RESULT 7

hypothetical protein F27B13.40 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Apr-2003  
 C:Accession: T08541  
 R:Bevan, M.; Zimmermann, W.; Gruenisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Maye  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16442  
 A:Accession: T08541  
 A:Molecule type: DNA  
 A:Residues: 1-526 <BEV>  
 A:Cross-references: EMBL:AL050352; GSPDB:GN00062; ATSP:F27B13.40  
 A:Experimental source: cultivar Columbia; BAC clone F27B13

C:Genetics:  
 A:Gene: ATSP:F27B13.40  
 A:Map position: 4  
 A:Introns: 421/3  
 C:Superfamily: patatin

Query Match 1.0%; Score 8; DB 2; Length 526;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 TPSSSFRS 124  
 |||||



Db 109 TPSSFRS 116

RESULT 8

JC7331  
gamma-glutamyltransferase (EC 2.3.2.2) homolog - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 15-Sep-2000  
C:Accession: JC7331  
R:Yamaguchi, T.; Takei, N.; Araki, K.; Ishii, K.; Nagano, T.; Ichikawa, T.; Kumanishi, T.  
J. Biochem. 128, 101-106, 2000  
A:Title: Molecular characterization of a novel gamma-glutamyl transpeptidase homologue f  
A:Reference number: JC7331  
A:Accession: JC7331  
A:Molecule type: mRNA  
A:Residues: 1-641 <YAM>  
A:Cross-references: GB:AF244973  
A:Experimental source: brain  
C:Comment: This enzyme is a membrane-bound glycoenzyme that is involved in the metabolism  
glutathione.  
C:Genetics:  
A:Gene: gtpb  
C:Keywords: aminoacyltransferase; brain; glutathione; glycolysis; membrane bound

Query Match 1.0%; Score 8; DB 2; Length 641;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 DMLSKVE 179  
|||||||  
Db 415 DMLSKVE 422

RESULT 9

T08827  
hypothetical protein cotel - human  
C:Species: Homo sapiens (man)  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T08827  
R:Winfield, S.L.; Tayebi, N.; Martin, B.M.; Ginns, E.I.; Sidransky, E.  
Genome Res. 7, 1020-1026, 1997  
A:Title: Identification of three additional genes contiguous to the glucocerebrosidase 1  
A:Reference number: Z16482; MUID:97474796; PMID:9331372  
A:Accession: T08827  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-669 <WIN>  
A:Cross-references: EMBL:AF023268; NID:g2564910; PIDN:AAC51822.1; PID:g2564916  
C:Genetics:  
A:Gene: cotel  
A:Map position: 1  
A:Introns: 75/3; 94/3; 131/3; 171/3; 207/3; 266/2; 299/3; 323/1; 505/1; 528/1; 612/3

Query Match 1.0%; Score 8; DB 2; Length 669;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 SSGSLTPP 111  
|||||||  
Db 467 SSGSLTPP 474

RESULT 10

VTVAA2  
viscotoxin A2 - European mistletoe  
C:Species: Viscum album (European mistletoe)  
C>Date: 13-Jul-1981 #sequence\_revision 08-Oct-1981 #text\_change 04-Oct-1996  
C:Accession: A90005; A01799  
R:Olson, T.; Samuelsson, G.  
Acta Chem. Scand. 26, 585-595, 1972  
A:Title: The amino acid sequence of viscotoxin A2 from the European mistletoe (Viscum al  
A:Reference number: A90005; MUID:72211843; PMID:5035954  
A:Accession: A90005

A:Molecule type: protein  
A:Residues: 1-46 <OLS>  
R:Olson, T.; Samuelsson, G.  
Acta Pharm. Suec. 11, 381-386, 1974  
A:Title: The disulphide bonds of viscotoxin A2 from the European mistletoe (Viscum album  
A:Reference number: A90013; MUID:75015879; PMID:4607177  
C:Contents: annotation; disulfide bonds  
C:Superfamily: viscotoxin  
C:Keywords: toxin  
F:3-40.4-32/Disulfide bonds: #status predicted  
F:16-26/Disulfide bonds: #status experimental

Query Match 0.9%; Score 7; DB 1; Length 46;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 NTCRFGG 18  
|||||||  
Db 14 NTCRFGG 20

RESULT 11

VFFD1T  
phoratoxin - California mistletoe  
C:Species: Phoradendron tomentosum (California mistletoe)  
C>Date: 22-Jun-1981 #sequence\_revision 22-Jun-1981 #text\_change 15-Oct-1996  
C:Accession: A01802  
R:Mellstrand, S.T.; Samuelsson, G.  
Acta Pharm. Suec. 11, 347-360, 1974  
A:Title: Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. macro  
A:Reference number: A90011; MUID:75013484; PMID:4415051  
A:Note: P. t. subsp. macrophyllum  
A:Accession: A01802

A:Molecule type: protein  
A:Residues: 1-46 <MEL>  
A:Note: the carboxyl end is blocked but is not amidated  
R:Mellstrand, S.T.; Samuelsson, G.  
Acta Pharm. Suec. 11, 367-374, 1974  
A:Title: Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. m  
A:Reference number: A90012; MUID:75013480; PMID:4606908  
C:Contents: annotation; disulfide bonds  
C:Comment: Phoratoxin, the single toxin of an American species of mistletoe, is related  
C:Superfamily: viscotoxin  
C:Keywords: blocked carboxyl end; toxin  
F:3-40.4-32/Disulfide bonds: #status predicted  
F:16-26/Disulfide bonds: #status experimental  
F:46/Modified site: blocked carboxyl end (His) #status experimental

Query Match 0.9%; Score 7; DB 1; Length 46;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 NTCRFGG 18  
|||||||  
Db 14 NTCRFGG 20

RESULT 12

S76642  
hypothetical protein - Synecocystis sp. (strain PCC 6803)  
C:Species: Synecocystis sp.  
A:Variety: PCC 6803  
C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S76642  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
s.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S76642  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A;Residues: 1-52 <KAN>  
 A;Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BA010586.1; PID:d101123  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C;Genetics:  
 A;Start codon: GTG

Query Match 0.9%; Score 7; DB 2; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 NIKRKTFF 169  
 |||||  
 DB 17 NIKRKTFF 23

## RESULT 13

AB1046  
 SUGB protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18)  
 C;Species: Salmonella enterica subsp. enterica serovar typhi  
 A;Note: this species has also been called Salmonella typhi  
 C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C;Accession: AB1046  
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001.  
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar  
 A;Reference number: AB0502; MUID:21534947; PMID:11677608  
 A;Accession: AB1046  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-105 <PAR>  
 A;Cross-references: GB:AL513382; PIDN:CAD06818.1; PID:g16505468; GSPDB:GN00176  
 C;Genetics:  
 A;Gene: sugE  
 C;Superfamily: sugE protein

Query Match 0.9%; Score 7; DB 2; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 GESASPA 734  
 |||||  
 DB 79 GESASPA 85

## RESULT 14

F97132  
 hypothetical protein CAC1885 [imported] - Clostridium acetobutylicum  
 C;Species: Clostridium acetobutylicum  
 C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C;Accession: F97132  
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A;Reference number: A96900; MUID:21359325; PMID:21359325  
 A;Accession: F97132  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-109 <KUR>  
 A;Cross-references: GB:AB001437; PIDN:AAK79849.1; PID:g15024865; GSPDB:GN00168  
 A;Experimental source: Clostridium acetobutylicum ATCC824  
 C;Genetics:  
 A;Gene: CAC1885

Query Match 0.9%; Score 7; DB 2; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 ANEEITE 740  
 |||||

Db 64 ANEEITE 70

## RESULT 15

H90626  
 NADH dehydrogenase chain 3 ND3 [imported] - Eudromia elegans mitochondrion  
 C;Species: Mitochondrion Eudromia elegans  
 C;Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 03-Aug-2001  
 C;Accession: H90626  
 R;Haddrath, O.; Baker, A.J.  
 Proc. R. Soc. Lond. B Biol. Sci. 268, 939-945, 2001  
 A;Title: Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetic  
 A;Reference number: A99613; MUID:21263106; PMID:11370967  
 A;Accession: H90626  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-116 <KUR>  
 A;Cross-references: GB:NC\_002772; NID:g14141811; PIDN:NP\_115280.1; GSPDB:GN00163  
 C;Genetics:  
 A;Gene: ND3  
 A;Genome: mitochondrion  
 A;Genetic code: SGC1  
 C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3  
 C;Keywords: mitochondrion

Query Match 0.9%; Score 7; DB 2; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSILTSS 90  
 |||||  
 DB 10 LSILTSS 16

Search completed: August 25, 2004, 18:15:33  
 Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:14:54 ; Search time 137 Seconds

(without alignments)  
1782.041 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 776

Sequence: 1 MTGIAAASFNTCRFGCGG.....KALETDSVSGVSKQKKQL 776

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1297172 seqs, 314612898 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1297172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCT05\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	776	100.0	776	US-09-874-162A-8	Sequence 8, Appli
2	647	83.4	739	US-09-874-162A-5	Sequence 5, Appli
3	283	36.5	289	US-09-764-864-911	Sequence 911, App
4	186	24.0	388	US-09-764-864-931	Sequence 931, App
5	161	20.7	292	US-09-764-864-1369	Sequence 1369, App
6	131	16.9	175	US-09-764-864-1355	Sequence 1355, App
7	128	16.5	243	US-09-874-162A-2	Sequence 2, Appli
8	128	16.5	243	US-10-144-198-16	Sequence 16, Appli
9	128	16.5	243	US-10-104-047-2404	Sequence 2404, Ap
10	109	14.0	278	US-09-764-864-1334	Sequence 1334, Ap
11	93	12.0	143	US-09-764-864-884	Sequence 884, App
12	67	8.6	69	US-10-029-386-28495	Sequence 28495, A
13	56	7.2	171	US-10-276-774-2060	Sequence 2060, Ap
14	8	1.0	25	US-09-864-761-42997	Sequence 42997, A
15	8	1.0	150	US-10-437-963-122951	Sequence 122951,

16	8	1.0	150	16	US-10-437-963-147736	Sequence 147736,
17	8	1.0	151	12	US-10-424-599-170262	Sequence 170262,
18	8	1.0	158	12	US-10-424-599-170263	Sequence 170263,
19	8	1.0	283	12	US-10-425-114-71700	Sequence 71700, A
20	8	1.0	285	12	US-10-425-114-51976	Sequence 51976, A
21	8	1.0	297	9	US-09-738-626-5528	Sequence 5528, Ap
22	8	1.0	324	12	US-10-425-114-37645	Sequence 37645, A
23	8	1.0	372	12	US-10-424-599-260059	Sequence 260059,
24	8	1.0	573	16	US-10-287-092-34	Sequence 34, Appl
25	8	1.0	662	9	US-09-924-358-11	Sequence 11, Appl
26	8	1.0	662	15	US-10-410-764-11	Sequence 11, Appl
27	8	1.0	669	10	US-09-790-852-7	Sequence 7, Appli
28	8	1.0	673	16	US-10-437-963-123277	Sequence 123277,
29	8	1.0	1022	14	US-10-301-822-163	Sequence 163, App
30	8	1.0	1022	16	US-10-648-593-244	Sequence 244, App
31	8	1.0	1093	16	US-10-275-998-1	Sequence 1, Appli
32	7	0.9	10	10	US-09-572-404B-3143	Sequence 3143, Ap
33	7	0.9	10	10	US-09-572-404B-3145	Sequence 3145, Ap
34	7	0.9	51	16	US-10-437-963-162817	Sequence 162817,
35	7	0.9	54	16	US-10-437-963-161985	Sequence 161985,
36	7	0.9	59	12	US-10-424-599-163032	Sequence 163032,
37	7	0.9	60	10	US-09-764-891-3469	Sequence 3469, Ap
38	7	0.9	63	12	US-10-424-599-155981	Sequence 155981,
39	7	0.9	66	12	US-10-424-599-259111	Sequence 259111,
40	7	0.9	69	16	US-10-437-963-135864	Sequence 135864,
41	7	0.9	70	16	US-10-437-963-147556	Sequence 147556,
42	7	0.9	74	12	US-10-450-875-6	Sequence 6, Appli
43	7	0.9	89	12	US-10-424-599-180215	Sequence 180215,
44	7	0.9	90	12	US-10-424-599-255393	Sequence 255393,
45	7	0.9	94	10	US-09-764-872-312	Sequence 312, App

ALIGNMENTS

RESULT 1

US-09-874-162A-8  
; Sequence 8, Application US/09874162A  
; Patent No. US20020155452A1  
; GENERAL INFORMATION:  
; APPLICANT: Koontz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JJA21 GENES IN  
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874,162A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,093  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 776  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-162A-8

Query Match	100.0%	Score 776;	DB 9;	Length 776;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	776;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
Qy	1	MTGIAAASFNTCRFGCGGLHPTTADLIEHEDNHDTPRVLEKQELQQPTTVALSY	60	
Db	1	MTGIAAASFNTCRFGCGGLHPTTADLIEHEDNHDTPRVLEKQELQQPTTVALSY	60	
Qy	61	INFMTDAARREOESLKKIQPKLSLTLSSVSRGVNSTPPRHSSGSLTPVPPTPTSS	120	
Db	61	INFMTDAARREOESLKKIQPKLSLTLSSVSRGVNSTPPRHSSGSLTPVPPTPTSS	120	
Qy	121	SFSSSTPTPTQYRFLRNLNLIAPFLHRTLTYMSHRNSTNKKTKFKVDDMLSKVEK	180	
Db	121	SFSSSTPTPTQYRFLRNLNLIAPFLHRTLTYMSHRNSTNKKTKFKVDDMLSKVEK	180	

181 MKGEQESHLSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRDXVSCPI 240  
181 MKGEQESHLSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRDXVSCPI 240  
241 ROVPTGKKQVPLIPDLNOKGPNPSSLAVSSNEPEPNSHVMKYSLLFRVTRPGRREFN 300  
241 ROVPTGKKQVPLIPDLNOKGPNPSSLAVSSNEPEPNSHVMKYSLLFRVTRPGRREFN 300  
301 GMINGETNENIDVNEELPARKRNRDEGKTFVAQMTVFDKRNRLQLLDGGEYVAMQEME 360  
301 GMINGETNENIDVNEELPARKRNRDEGKTFVAQMTVFDKRNRLQLLDGGEYVAMQEME 360  
361 ECPISKGAATWETILDGKRLPPFTFSOGPTLQTLRWGTGTNDKSTAPIAKPLATRNSE 420  
361 ECPISKGAATWETILDGKRLPPFTFSOGPTLQTLRWGTGTNDKSTAPIAKPLATRNSE 420  
421 SLHQENKSGSVKPTOTIAVKSLLTDLQTRKEKDTNENRQKLRIFYQFLYNNNTROOTE 480  
421 SLHQENKSGSVKPTOTIAVKSLLTDLQTRKEKDTNENRQKLRIFYQFLYNNNTROOTE 480  
481 ARDDLHCPWCTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKIDKAVTKLRE 720  
481 ARDDLHCPWCTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKIDKAVTKLRE 720  
541 QDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDDGEVEQOITYSSGHN 600  
541 QDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDDGEVEQOITYSSGHN 600  
601 LYFHSDDTCLPLRPOQMEVDSDEKDPWLREKTTIQIEEFSNVNEGEKVMKLNHLVHK 660  
601 LYFHSDDTCLPLRPOQMEVDSDEKDPWLREKTTIQIEEFSNVNEGEKVMKLNHLVHK 660  
661 HGFIADNOMHACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLRE 720  
661 HGFIADNOMHACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLRE 720  
721 MOOKLEKESASPAEETEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
721 MOOKLEKESASPAEETEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776

## RESULT 2

US-09-874-162A-5  
; Sequence 5, Application US/09874162A  
; Patent No. US2002015542A1  
; GENERAL INFORMATION:  
; APPLICANT: Koontz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZA1 GENES IN  
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874,162A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,093  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 739  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-162A-5

Query Match 83.4%; Score 647; DB 9; Length 739;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

130 PTQIYRFLRTRNLIAPIFLHRTLYMHSRNRTRNKRTFKVDDMLSKVEMKGEQESH 189  
93 PTQIYRFLRTRNLIAPIFLHRTLYMHSRNRTRNKRTFKVDDMLSKVEMKGEQESH 152  
190 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRDXVSCPIRQVPTGKKQ 249

153 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLEVLVVKVCHKKRDXVSCPIRQVPTGKKQ 212  
250 VPLIPDLNOKGPNPSSLAVSSNEPEPNSHVMKYSLLFRVTRPGRREFNMGINGETNE 309  
213 VPLIPDLNOKGPNPSSLAVSSNEPEPNSHVMKYSLLFRVTRPGRREFNMGINGETNE 272  
310 NIDVNEELPARKRNRDEGKTFVAQMTVFDKRNRLQLLDGGEYVAMQEMECPISKRA 369  
273 NIDVNEELPARKRNRDEGKTFVAQMTVFDKRNRLQLLDGGEYVAMQEMECPISKRA 332  
370 TWETILDGKRLPPFTFSOGPTLQTLRWGTGTNDKSTAPIAKPLATRNSSLHGENKPG 429  
333 TWETILDGKRLPPFTFSOGPTLQTLRWGTGTNDKSTAPIAKPLATRNSSLHGENKPG 392  
430 SVKPTQTTIAVKSLLTDLQTRKEKDTNENRQKLRIFYQFLYNNNTROOTEARDLHCPW 489  
393 SVKPTQTTIAVKSLLTDLQTRKEKDTNENRQKLRIFYQFLYNNNTROOTEARDLHCPW 452  
490 CTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKIDKAVTKLREKQKLEGE 549  
453 CTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKIDKAVTKLREKQKLEGE 512  
550 AFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDDGEVEQOITYSSGHNRLYFHSDDTCL 609  
513 AFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDDGEVEQOITYSSGHNRLYFHSDDTCL 572  
610 PLRPOQMEVDSDEKDPWLREKTTIQIEEFSNVNEGEKVMKLNHLVHKHGFADNOM 669  
573 PLRPOQMEVDSDEKDPWLREKTTIQIEEFSNVNEGEKVMKLNHLVHKHGFADNOM 632  
670 NEACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLREKQKLEGE 729  
633 NEACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLREKQKLEGE 692  
730 SASPANEETEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
693 SASPANEETEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 739

## RESULT 3

US-09-764-864-911  
; Sequence 911, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 911  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-911

Query Match 36.5%; Score 283; DB 9; Length 289;  
Best Local Similarity 100.0%; Pred. No. 2,1e-270;  
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

494 CRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKIDKAVTKLREKQKLEGE 553  
7 CRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKIDKAVTKLREKQKLEGE 66  
554 NGPVKRTPTIHLVCRPKRTKASMEFLESEDDGEVEQOITYSSGHNRLYFHSDDTCLPLRP 613  
67 NGPVKRTPTIHLVCRPKRTKASMEFLESEDDGEVEQOITYSSGHNRLYFHSDDTCLPLRP 126  
614 QMEVDSDEKDPWLREKTTIQIEEFSNVNEGEKVMKLNHLVHKHGFADNOMHAC 673  
127 QMEVDSDEKDPWLREKTTIQIEEFSNVNEGEKVMKLNHLVHKHGFADNOMHAC 186

QY 674 MLFVNYGQKIKNLGRNFMHLVSMDFNLSIMSIDKAVTKLRMOOKLEKESASP 733  
Db 187 MLFVNYGQKIKNLGRNFMHLVSMDFNLSIMSIDKAVTKLRMOOKLEKESASP 246  
QY 734 ANEEITEONGTANGFSEINSKEKALTDVSVGSKQSKOKL 776  
Db 247 ANEEITEONGTANGFSEINSKEKALTDVSVGSKQSKOKL 289

## RESULT 4

US-09-764-864-931  
; Sequence 931, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 931  
; LENGTH: 388  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-931

Query Match 24.0%; Score 186; DB 9; Length 388;  
Best Local Similarity 99.5%; Pred. No. 1.6e-174;  
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 389 GPTLOFTLWTGETNDKSTAPIAKPLATNSLSHQENKPGSVKPTQTIIVAKESLTTDLQ 448  
Db 1 GPTLOFTLWTGETNDKSTAPIAKPLATNSLSHQENKPGSVKPTQTIIVAKESLTTDLQ 60  
QY 449 TRKEKDTNENRQKRIFYQFLYNNNTROOTEARDLHCPWCTLNCRKLYSLHLKHLCH 508  
Db 61 TRKEKDTNENRQKRIFYQFLYNNNTROOTEARDLHCPWCTLNCRKLYSLHLKHLCH 120  
QY 509 SRPIFYVYHPKGAIDVINSICYGVSAGNPQDIHRQPGFAPSRNGPVKPTPIHILVC 568  
Db 121 SRPIFYVYHPKGAIDVINSICYGVSAGNPQDIHRQPGFAPSRNGPVKPTPIHILVC 180  
QY 569 RPKRTKASSEFSEDEGEVEQRTYSSGHNLRYFHSHTCLPLRQEMEVDSDEKDPW 628  
Db 181 RPKRTKASSEFSEDEGEVEQRTYSSGHNLRYFHSHTCLPLRQEMEVDSDEKDPW 240  
QY 629 LREKTIQIEFSDVNEGEKYMKNLWLVKMGHFIADNQMNHACMLFVNYGQKIKN 688  
Db 241 LREKTIQIEFSDVNEGEKYMKNLWLVKMGHFIADNQMNHACMLFVNYGQKIKN 300  
QY 689 LCRNFMHLVSMDFNLSIMSIDKAVTKLRMOOKLEKESASPANEEITEONGTANG 748  
Db 301 LCRNFMHLVSMDFNLSIMSIDKAVTKLRMOOKLEKESASPANEEITEONGTANG 360  
QY 749 FSEINSKEKALTDVSVGSKQSKOKL 776  
Db 361 FSEINSKEKALTDVSVGSKQSKOKL 388

## RESULT 5

US-09-764-864-1369  
; Sequence 1369, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1369  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (9)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (16)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (18)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (31)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (46)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (121)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (283)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-864-1369

Query Match 20.7%; Score 161; DB 9; Length 292;  
Best Local Similarity 100.0%; Pred. No. 5.9e-150; Indels 0; Gaps 0;  
Matches 161; Conservative 0; Mismatches 0;  
QY 354 VAMQEMEECPISKGRATWETILDGKRLPPETFSQGPFTLQTLRWGTETNDKSTAPIAKP 413  
Db 122 VAMQEMEECPISKGRATWETILDGKRLPPETFSQGPFTLQTLRWGTETNDKSTAPIAKP 181  
QY 414 LATNSLSHQENKPGSVKPTQTIIVAKESLTTDLQTRKEKDTNENRQKRIFYQFLYNN 473  
Db 182 LATNSLSHQENKPGSVKPTQTIIVAKESLTTDLQTRKEKDTNENRQKRIFYQFLYNN 241  
QY 474 NTRQOTEARDLHCPWCTLNCRKLYSLHLKHLCHSRFIEN 514  
Db 242 NTRQOTEARDLHCPWCTLNCRKLYSLHLKHLCHSRFIEN 282

## RESULT 6

US-09-764-864-1355  
; Sequence 1355, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1355  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (136)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (166)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-864-1355

Query Match 16.5%; Score 131; DB 9; Length 175;  
Best Local Similarity 100.0%; Pred. No. 1.9e-120; Indels 0; Gaps 0;  
Matches 131; Conservative 0; Mismatches 0;  
QY 470 LYNNTRQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIYVYHPKGARIDVSN 529  
Db LYNNTRQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIYVYHPKGARIDVSN 64  
QY 530 EYDGSYAGNPDIHQPGAFSRNGPVKRTPIITHILVCRPKTKASMEFESEDEGE 589  
Db EYDGSYAGNPDIHQPGAFSRNGPVKRTPIITHILVCRPKTKASMEFESEDEGE 124  
QY 590 QORTYSSGHN 600  
Db QORTYSSGHN 135

RESULT 7  
US-09-874-162A-2  
; Sequence 2, Application US/09874162A  
; Patent No. US20020155452A1  
; GENERAL INFORMATION:  
; APPLICANT: Koonetz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JUJAZ1 GENES IN  
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874,162A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,093  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-162A-2

Query Match 16.5%; Score 128; DB 9; Length 243;  
Best Local Similarity 100.0%; Pred. No. 1.9e-117; Indels 0; Gaps 0;  
Matches 128; Conservative 0; Mismatches 0;  
QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDPVLEKQELQOPTYVALSY 60  
Db 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDPVLEKQELQOPTYVALSY 60  
QY 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSGNVSTPPRHSSGSLTPPVTPITPSS 120  
Db 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSGNVSTPPRHSSGSLTPPVTPITPSS 120  
QY 121 SFRSSTPT 128  
Db 121 SFRSSTPT 128

RESULT 8  
US-10-144-198-16  
; Sequence 16, Application US/10144198  
; Publication No. US20030219748A1  
; GENERAL INFORMATION:  
; APPLICANT: OriGene Technologies Inc  
; TITLE OF INVENTION: Regulated Prostate Cance Genes  
; FILE REFERENCE: 9U 105 R1  
; CURRENT APPLICATION NUMBER: US/10/144,198  
; CURRENT FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 16  
; LENGTH: 243

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-144-198-16  
Query Match 16.5%; Score 128; DB 15; Length 243;  
Best Local Similarity 100.0%; Pred. No. 1.9e-117; Indels 0; Gaps 0;  
Matches 128; Conservative 0; Mismatches 0;  
QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDPVLEKQELQOPTYVALSY 60  
Db 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDPVLEKQELQOPTYVALSY 60  
QY 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSGNVSTPPRHSSGSLTPPVTPITPSS 120  
Db 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSGNVSTPPRHSSGSLTPPVTPITPSS 120  
QY 121 SFRSSTPT 128  
Db 121 SFRSSTPT 128

RESULT 9  
US-10-104-047-2404  
; Sequence 2404, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2404  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2404

Query Match 16.5%; Score 128; DB 15; Length 243;  
Best Local Similarity 100.0%; Pred. No. 1.9e-117; Indels 0; Gaps 0;  
Matches 128; Conservative 0; Mismatches 0;  
QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDPVLEKQELQOPTYVALSY 60  
Db 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDPVLEKQELQOPTYVALSY 60  
QY 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSGNVSTPPRHSSGSLTPPVTPITPSS 120  
Db 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSGNVSTPPRHSSGSLTPPVTPITPSS 120  
QY 121 SFRSSTPT 128  
Db 121 SFRSSTPT 128

RESULT 10  
US-09-764-864-1334  
; Sequence 1334, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT23  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1334  
; LENGTH: 278

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1334

Query Match
Best Local Similarity 14.0%; Score 109; DB 9; Length 278;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDRVLEKQELQOPTVVALSY 60
Db 36 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDRVLEKQELQOPTVVALSY 95
QY 61 INRFMTDAARRQESLKKKIQKLSLTLSVVSRGNVSTPPRHSSGSLT 109
Db 96 INRFMTDAARRQESLKKKIQKLSLTLSVVSRGNVSTPPRHSSGSLT 144

RESULT 11
US-09-764-864-884
; Sequence 884, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 884
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-884

Query Match
Best Local Similarity 12.0%; Score 93; DB 9; Length 143;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDRVLEKQELQOPTVVALSY 60
Db 51 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDRVLEKQELQOPTVVALSY 110
QY 61 INRFMTDAARRQESLKKKIQKLSLTLSVVSRGNVSTPPRHSSGSLT 143
Db 111 INRFMTDAARRQESLKKKIQKLSLTLSVVSRGNVSTPPRHSSGSLT 143

RESULT 12
US-10-029-386-28495
; Sequence 28495, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28495
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; OTHER INFORMATION: MAP TO CHRI17.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P23477, EVALUATE 2.70e+00
US-10-029-386-28495

Query Match
Best Local Similarity 8.6%; Score 67; DB 14; Length 69;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 RPKETKASMEFFLESEDEGEVQORTYSSGHNRLYFHSDTCLPLRPQMEYDSEDEKDPW 628
Db 1 RPKETKASMEFFLESEDEGEVQORTYSSGHNRLYFHSDTCLPLRPQMEYDSEDEKDPW 60
QY 629 LREXITIT 635
Db 61 LREXITIT 67

RESULT 13
US-10-276-774-2060
; Sequence 2060, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2060
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2060

Query Match
Best Local Similarity 7.2%; Score 56; DB 12; Length 171;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 QESLKKKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTTPVTPBTSSSSPSSSTPT 128
Db 1 QESLKKKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTTPVTPBTSSSSPSSSTPT 56

RESULT 14
US-09-864-761-42997
; Sequence 42997, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366

```

```

; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42997
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012111.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
US-09-864-761-42997

```

```

Query Match          1.0%; Score 8; DB 9; Length 25;
Best Local Similarity 100.0%; Pred.No. 7.8;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      110 PPVTPPIT 117
          |||||
DB      13 PPVTPPIT 20

```

```

RESULT 15
US-10-437-963-122951
; Sequence 122951, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963

```

```

; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122951
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MPT4530_25832C.1.pep
US-10-437-963-122951

Query Match          1.0%; Score 8; DB 16; Length 150;
Best Local Similarity 100.0%; Pred.No. 38;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      408 APIAKPLA 415
          |||||
DB      17 APIAKPLA 24

Search completed: August 25, 2004, 18:26:48
Job time : 140 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:09:41 ; Search time 33 Seconds  
(without alignments)  
1213.993 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 776  
Sequence: 1 MTGIAAGFFSNTCRFGCG.....KALETDSVSGVSKOSKQKL 776

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	1.3	64	4	US-09-543-681A-6424
2	9	1.2	96	4	US-09-328-352-4313
3	7	0.9	139	4	US-09-770-834-13
4	7	0.9	142	4	US-09-328-352-4320
5	7	0.9	151	4	US-09-621-976-4980
6	7	0.9	155	4	US-09-489-039A-14046
7	7	0.9	174	4	US-09-328-352-7878
8	7	0.9	178	4	US-09-328-352-4296
9	7	0.9	196	4	US-09-842-164A-2
10	7	0.9	206	4	US-09-252-991A-31591
11	7	0.9	247	4	US-09-134-001C-5610
12	7	0.9	279	4	US-09-107-532A-6807
13	7	0.9	294	3	US-08-973-334-2
14	7	0.9	294	3	US-09-563-869A-2
15	7	0.9	294	3	US-08-549-489-2
16	7	0.9	300	4	US-09-252-991A-17565
17	7	0.9	320	2	US-08-579-940-8
18	7	0.9	324	4	US-09-010-147B-16
19	7	0.9	324	4	US-09-498-520A-10
20	7	0.9	332	4	US-09-134-000C-4988
21	7	0.9	333	4	US-09-540-236-2603
22	7	0.9	368	4	US-08-630-915A-20
23	7	0.9	370	4	US-09-107-532A-5268
24	7	0.9	384	4	US-09-198-452A-1111
25	7	0.9	402	4	US-09-721-870-46
26	7	0.9	493	4	US-09-252-991A-27576
27	7	0.9	519	3	US-08-996-338-21

28	7	0.9	519	4	US-09-556-972-21	Sequence 21, Appl
29	7	0.9	526	4	US-09-198-452A-347	Sequence 347, App
30	7	0.9	534	3	US-08-875-223-8	Sequence 8, Appli
31	7	0.9	537	1	US-08-604-333-4	Sequence 4, Appli
32	7	0.9	537	3	US-09-110-618-4	Sequence 4, Appli
33	7	0.9	537	4	US-09-173-151A-29	Sequence 29, Appl
34	7	0.9	537	4	US-09-578-178-4	Sequence 4, Appli
35	7	0.9	537	4	US-09-577-806-4	Sequence 4, Appli
36	7	0.9	537	4	US-09-621-502-8	Sequence 8, Appli
37	7	0.9	566	3	US-09-073-297-5	Sequence 5, Appli
38	7	0.9	570	3	US-08-991-944-2	Sequence 2, Appli
39	7	0.9	570	3	US-08-991-944-4	Sequence 4, Appli
40	7	0.9	570	4	US-09-173-151A-21	Sequence 21, Appl
41	7	0.9	570	4	US-09-173-151A-36	Sequence 36, Appl
42	7	0.9	602	4	US-09-489-039A-7207	Sequence 7207, Ap
43	7	0.9	690	4	US-09-252-991A-32350	Sequence 32350, A
44	7	0.9	887	4	US-09-077-940A-2	Sequence 2, Appli
45	7	0.9	888	4	US-09-077-940A-4	Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-09-543-681A-6424  
; Sequence 6424, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 6424  
; LENGTH: 64  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-6424

Query Match 1.3%; Score 10; DB 4; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.02; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	85	SLTLSSSVSR	94
Db	48	SLTLSSSVSR	57

RESULT 2  
US-09-328-352-4313  
; Sequence 4313, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4313  
; LENGTH: 96  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4313

Query Match 1.2%; Score 9; DB 4; Length 96;  
Best Local Similarity 100.0%; Pred. No. 0.44; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 TPPVTPPIT 117  
|||||  
Db 55 TPPVTPPIT 63

## RESULT 3

US-09-770-834-13  
; Sequence 13, Application US/09770834  
; Patent No. 6684162

## ; GENERAL INFORMATION:

; APPLICANT: Parris, Kevin

; APPLICANT: Somers, William

; APPLICANT: Tam, Amy

; APPLICANT: Lin, Laura

; APPLICANT: Stahl, Mark

; APPLICANT: Powers, Robert

; APPLICANT: Xu, Guan-Yi

; TITLE OF INVENTION: CRYSTAL STRUCTURE OF ACPS/ACP COMPLEX, SOLUTION STRUCTURE  
; FILE OF INVENTION: OF B. SUBTILIS ACP, AND USES THEREOF

; FILE REFERENCE: 2368/14

; CURRENT APPLICATION NUMBER: US/09/770,834

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/202,466

; PRIOR FILING DATE: 2000-05-08

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patent version 3.0

; SEQ ID NO 13

; LENGTH: 139

; TYPE: PRT

; ORGANISM: Bradyrhizobium sp.

US-09-770-834-13

## Query Match

Best Local Similarity 0.9%; Score 7; DB 4; Length 139;

; Sequence 4320, Application US/09328352

; Patent No. 6562958

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 522 ARIDVSI 528

Db 112 ARIDVSI 118

## RESULT 4

US-09-328-352-4320

; Sequence 4320, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 4320

; LENGTH: 142

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-4320

## Query Match

Best Local Similarity 0.9%; Score 7; DB 4; Length 142;

; Sequence 4320, Application US/09328352

; Patent No. 6562958

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 PPVTPPI 116

Db 37 PPVTPPI 43

## RESULT 5

US-09-621-976-4980

; Sequence 4980, Application US/09621976

; Patent No. 6639063

## ; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 4980

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-621-976-4980

## Query Match

Best Local Similarity 0.9%; Score 7; DB 4; Length 151;

; Sequence 14046, Application US/09489039A

; Patent No. 6610836

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 INECYDG 534

Db 55 INECYDG 61

## RESULT 6

US-09-489-039A-14046

; Sequence 14046, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 14046

; LENGTH: 155

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-14046

## Query Match

Best Local Similarity 0.9%; Score 7; DB 4; Length 155;

; Sequence 14046, Application US/09489039A

; Patent No. 6610836

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 GARIDVS 527

Db 57 GARIDVS 63

## RESULT 7

US-09-328-352-7878

; Sequence 7878, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 7878

; LENGTH: 174

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-7878

## Query Match

Best Local Similarity 0.9%; Score 7; DB 4; Length 174;

```
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 584 EDGEVEQ 590
Db 57 EDGEVEQ 63

RESULT 8
US-09-328-352-4296
; Sequence 4296, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4296
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4296

Query Match 0.9%; Score 7; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 PPVTPI 116
Db 138 PPVTPI 144

RESULT 9
US-09-842-164A-2
; Sequence 2, Application US/09842164A
; Patent No. 6544754
; GENERAL INFORMATION:
; APPLICANT: INOUE, SATOSHI
; TITLE OF INVENTION: LUCIFERASE AND PHOTOPROTEIN
; FILE REFERENCE: 206497US0
; CURRENT APPLICATION NUMBER: US/09/842,164A
; CURRENT FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: JP 2000-125053
; PRIOR FILING DATE: 2000-04-26
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Oplophorus graciliorostris
US-09-842-164A-2

Query Match 0.9%; Score 7; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 SLLFRVT 292
Db 175 SLLFRVT 181

RESULT 10
US-09-252-991A-31591
; Sequence 31591, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
```

```
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31591
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31591

Query Match 0.9%; Score 7; DB 4; Length 206;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 SLSAHLQ 195
Db 91 SLSAHLQ 97

RESULT 11
US-09-134-001C-5610
; Sequence 5610, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5610
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5610

Query Match 0.9%; Score 7; DB 4; Length 247;
Best Local Similarity 100.0%; Pred. No. 11e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 LLXHLKL 506
Db 124 LLXHLKL 130

RESULT 12
US-09-107-532A-6807
; Sequence 6807, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
```

```
;
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6807:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...279
; SEQUENCE DESCRIPTION: SEQ ID NO: 6807:
US-09-107-532A-6807

Query Match 0.9%; Score 7; DB 4; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 RVLEKQE 49
DB 65 RVLEKQE 71

RESULT 13
US-08-973-334-2
; Sequence 2, Application US/08973334
; Patent No. 6261551
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: Recombinant Adenovirus and Adeno-
; TITLE OF INVENTION: Associated Virus, Cell Lines, and
; TITLE OF INVENTION: Methods of Production and Use
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Box 457, 321 No. 6261551ristown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release 1.0 Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,334
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,014
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:

;
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6807:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...279
; SEQUENCE DESCRIPTION: SEQ ID NO: 6807:
US-09-107-532A-6807

Query Match 0.9%; Score 7; DB 4; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 RVLEKQE 49
DB 65 RVLEKQE 71

RESULT 13
US-08-973-334-2
; Sequence 2, Application US/08973334
; Patent No. 6261551
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: Recombinant Adenovirus and Adeno-
; TITLE OF INVENTION: Associated Virus, Cell Lines, and
; TITLE OF INVENTION: Methods of Production and Use
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Box 457, 321 No. 6261551ristown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release 1.0 Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,334
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,014
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:

;
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6807:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...279
; SEQUENCE DESCRIPTION: SEQ ID NO: 6807:
US-09-107-532A-6807

Query Match 0.9%; Score 7; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 RLPPFET 385
DB 29 RLPPFET 35

RESULT 14
US-09-563-869A-2
; Sequence 2, Application US/09563869A
; Patent No. 6270996
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: Recombinant Adenovirus and Adeno-
; TITLE OF INVENTION: Associated Virus, Cell Lines, and
; TITLE OF INVENTION: Methods of Production and Use
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Box 457, 321 No. 6270996ristown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release 1.0 Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/563,869A
; FILING DATE: 03-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/973,334
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/549,489
; FILING DATE: 27-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVN012CIPUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
```

```

;
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-563-869A-2

Query Match          0.9%; Score 7; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 RLPPET 385
Db 29 RLPPET 35

RESULT 15
US-08-549-489-2
; Sequence 2, Application US/08549489
; Patent No. 6281010
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: No. 6281010el Adenovirus Gene Therapy Vehicle
; TITLE OF INVENTION: and Cell Line
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Howson and Howson
; STREET: Box 457, 321 No. 6281010ristown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,489
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,014
; FILING DATE: 08-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVFN013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-549-489-2

Query Match          0.9%; Score 7; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 RLPPET 385
Db 29 RLPPET 35
```

Search completed: August 25, 2004, 18:16:11  
Job time : 35 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:43:38 ; Search time 133 Seconds  
(without alignments)  
1840.918 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 4078  
Sequence: 1 MTGIAAASFFNTCRFGCG.....KALETDSVSGVSKSQKQL 776

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriaph.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244.5	6.0	626	10 Q84UX5	Q84ux5 zea mays (m
2	224	5.5	628	5 Q9VZ26	Q9vz26 drosophila
3	205.5	5.0	614	10 Q7XSM1	Q7xsm1 oryza sativ
4	177	4.3	1422	5 Q8IBL9	Q8ibl9 plasmodium
5	174	4.3	1693	5 Q86H41	Q86h41 dictyosteli
6	170.5	4.2	1396	11 Q810T3	Q810t3 mus musculu
7	166	4.1	1507	6 Q95151	Q95151 canis famil
8	163	4.0	1314	3 Q53904	Q53904 kluyveromyc
9	160.5	3.9	2847	5 Q9U4X0	Q9u4x0 plasmodium
10	159.5	3.9	1401	6 Q8HY4	Q8hy4 bos taurus
11	158	3.9	3471	6 Q8MKI9	Q8mki9 canis famil
12	157.5	3.9	5767	5 Q81525	Q81525 plasmodium
13	157	3.8	1011	2 Q9AHL0	Q9ahl0 borrelia bu
14	157	3.8	1163	5 Q81KP5	Q81kp5 plasmodium
15	157	3.8	1893	5 Q9W4G6	Q9w4g6 drosophila
16	156.5	3.8	1065	2 Q9AHK7	Q9ahk7 borrelia bu

17	156.5	3.8	1173	2 Q9AHK6	Q9ahk6 borrelia bu
18	155	3.8	1233	5 Q8IJ56	Q8ij56 plasmodium
19	154.5	3.8	960	5 Q8ICR3	Q8icr3 plasmodium
20	154	3.8	1065	2 Q9AHK8	Q9ahk8 borrelia bu
21	154	3.8	1271	5 Q25860	Q25860 plasmodium
22	153.5	3.8	3692	16 Q8CNU9	Q8cnu9 staphylococ
23	153	3.8	1777	5 Q8I3P4	Q8i3p4 plasmodium
24	152	3.7	1119	16 O51228	O51228 borrelia bu
25	152	3.7	1749	3 Q99337	Q99337 saccharomyc
26	151	3.7	2033	5 Q8IM18	Q8im18 plasmodium
27	149	3.7	1145	3 Q94043	Q94043 candida alb
28	149	3.7	1236	5 Q9GTX2	Q9gtx2 plasmodium
29	148	3.6	898	3 Q96U06	Q96u06 neurospora
30	147.5	3.6	1368	5 Q25661	Q25661 plasmodium
31	147.5	3.6	1790	3 Q07380	Q07380 saccharomyc
32	146	3.6	16215	5 Q9NFS3	Q9nfs3 drosophila
33	146	3.6	18074	5 Q9I7U4	Q9i7u4 drosophila
34	145	3.6	1838	5 Q86IV6	Q86iv6 dictyosteli
35	145	3.6	2843	5 Q963L5	Q963l5 dictyosteli
36	144.5	3.5	1206	16 Q895S7	Q895s7 clostridium
37	144.5	3.5	2279	5 Q8IJP6	Q8ijp6 plasmodium
38	144	3.5	753	10 Q9LNV2	Q9lnv2 arabidopsis
39	144	3.5	2110	5 Q8IBF2	Q8ibf2 plasmodium
40	143.5	3.5	1190	5 Q9VNS1	Q9vns1 drosophila
41	143.5	3.5	17382	5 Q95YM2	Q95ym2 procamburus
42	143	3.5	1383	5 Q8I659	Q8i659 plasmodium
43	143	3.5	1953	3 O13450	O13450 saccharomyc
44	142.5	3.5	1461	5 Q95Z59	Q95z59 plasmodium
45	142	3.5	1051	5 Q95QW6	Q95qw6 caenorhabdi

ALIGNMENTS

RESULT 1

Q84UX5	PRELIMINARY;	PRT;	626 AA.
ID	Q84UX5		
AC	Q84UX5		
DT	01-JUN-2003 (TREMBLrel. 24, Created)		
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	VEF family protein.		
GN	VEF101.		
OS	Zea mays (Maize).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	PACCAD clade; Panicoideae; Andropogoneae; Zea.		
OX	NCBI_TaxID=4577;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. B73;		
RA	Chandler V.L., Kaeppler S.M., Kaeppler H.F., Cone K.C.;		
RT	"Sequences from the Plant Chromatin Consortium (NSF Plant Genome		
RT	program Grant 9975930).";		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. B73;		
RA	Bergstrom D.E., Springer N.M., Schmitt L., Guthrie E., Sidorenko L.,		
RA	Selinger D., Kaeppler S.M., Cone K.C.;		
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY232824; AAC04022.1; -		
DR	InterPro; IPR007087; Znf C2H2.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.		
SQ	SEQUENCE 626 AA; 71401 MW; 604COD46B5A8718 CRC64;		
Query Match			
Best Local Similarity 18.1%; Pred. No. 3.7e-09;			
Matches 123; Conservative 110; Mismatches 251; Indels 197; Gaps 22;			
Qy	129 EPTQIVREFLTRNLAPIFLHRLITLTYMSHRNSRNTKRTKFKVDDMLSKVKMKGEQESH 188		
Db	53 KPVELYNIQRRANKNPLFIQRCLLYNIHARRKRIQ----- 89		

QY 189 SUSALHQLTFTGFFHNDKPSNSNEQNSV-TLEVLIVKYCHKRRKDVSPICRQVPT-- 245  
DB 90 -----ITISLSG-----STNTELOTHYVPLVLLARPTNSLSEGHSPYFRSRVC 136  
QY 246 -----GKKQ-----VPLIPDLNQTGKGFSLAVSNEPEPNSHNVKSYLLPR 290  
DB 137 LUTSSEHGNKNSAATIPDVKSLSRACNHDIIIFISCGV----- 180  
QY 291 VTRPGREFNGMNGETNENIDVNEELPARRKRNEDEGKTFAQMTVDFKRNRLQLLDG 350  
DB 181 -----GQSGNEDNCS-----GNHVEDSS-----LQMLEG 204  
QY 351 EYEVAMQEMEECPISKKATW---ETILDGKELPPFFTFSSQPTLOFTLRWTGETNDKST 407  
DB 205 -----KCSWGKIPTNLLAGSLSCVNLGLHIVELASKVTMPRS----- 243  
QY 408 APIAKPLATRNSESL-----HOENKPGSVKPTQTAVKESLTDL-----QTRKEKDPNE 458  
DB 244 --FLEPKTLEQDSCLTFCGSHKVDVAGSYKLOLCHSAQAGARDMSLSPYSSYSYNDVPPS 301  
QY 459 NRQKL-----RIFQFLYNNNTROOTEARDLHCPWCTLNCRKLYSLKHLKCHSRFI 512  
DB 302 SLSDIIRLSGNVLNFYKYNTMTQETVTEDFSCFFCYVRCGSKGLGCHLNSHDLFH 361  
QY 513 FNYVYHPKARIDVGIN-ECYDGSVAGNPQDIHQPGFAF-----SRNGPVK-RTP 561  
DB 362 YEFWISSEYQVNVSLKADARWTELPARGVDPRQT-FSYRSRPFKRRRSKNTWEKIRHV 420  
QY 562 ITHILVC-RPKRTKASMEFLESDG----- 586  
DB 421 HSHIMESGPEDEAGSDFNGENGTSVANASIDPAQSLGCSNLSPTVLQFGKTRKLS 480  
QY 587 EYEQORTYSSGHNRLYFHSDTCLPLRPQEM--EYDSDEKDPFWLREKTIQIIEFSDVN 644  
DB 481 ERSDFRNQLQKROFFHSHRAQMQLEQVSDSDSEVDDDIADFEDRMLDDFVDVT 540  
QY 645 EGEKEVMKLNHVMKHGFIADNQNMHACMLFVENVYQKIKK-NLCRNFMLHLVSMHDF 703  
DB 541 KDEKLIMEMNSFVRKQVLADGHPWACEAFSLQHGRLQLNALLGWGFFMKLWNH 600  
QY 704 NLISMSIDKAVTKLRENOQK 724  
DB 601 NILDARTWNTCNTVLIQLOEE 621

## RESULT 2

Q9V926 PRELIMINARY; PRT: 628 AA.  
AC Q9V926; Q960L0;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE CG12054-PA (LD46863p).  
CN CG12054  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazei R.G., Champe M., Pfeiffer B.D.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosiier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodsk A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
[2]  
RN SEQUENCE FROM N.A.  
RP Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gozalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE FROM N.A.  
RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RN SEQUENCE FROM N.A.  
RP FlyBase.  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
[6]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Berkeley.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C. Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.



[illegible]





```

Db      655 KCRLL-ILQNNIEMTIDIFVEETSYRNTENTENEGNQCTDAGRNTCNSESQSDSSKNDTV 713
QY      312 DYNEE---LPARKENREDGEKTFVAQMTVPDKRRRLQLLDGEYEV---AMQMEECPI 364
Db      714 YIHEENGSLPCIDQHNID-----LKFSQFMKEGNTQIKESLDLTCLEVMKAEETSHV 767
QY      365 SKKRATWETILDGKLPFPFEP---SQGPTLQFTLRWGETNDKSTAPIAKPLATRN-----418
Db      768 TWSNKQQTANTGQNIKDFDYFYLFSQFASRKNIKRVSRSLNKARSLLNOKWTEELNPF 827
QY      419 SESLHQNCKPG-SVKPTQTIIVAKESLTLTDLQTRKEKTPNENRQKLRIFYQFLYNNNTRQ 477
Db      828 SDLSNELLPGLDIIKTD---ISNHEVIENTERKDKITKESDLTGENTILLILOQRPESK 884
QY      478 QTEARDDHLCPWCTLNCRLKLYSLKHLKCHSRP-----IFNVYHPKGIARDIVSI 528
Db      885 IKKIKESAVLGFHTASGKKIETITKESLDKVNLFEEKEQDNSEITNFSH--RGAKMSKDR 942
QY      529 NECYDG-SYAGNPDIHQPGFAPFASRNGPVRKTPITH-ILVCRP-----KRT-----573
Db      943 BECKDGRACGTBEITTPYEETHSSLEKLVSEIAALRPRLSLDNLKYOTENLKI 1002
QY      574 -----KASMEFELESDGE-----VEQRTYSNGHR-LYPHSDTCLPLRPQMEVDS 622
Db      1003 SDHASQKVDVHENTEKETAKKPTMYTNGSTYSIAENSPLTFTQDT-----EEKFSVSEA 1056
QY      623 E--KDPKRLREKTIQIIEPSPD-----VNEGEKEVMKWLW-----NLHVHKGFADNQ 668
Db      1057 SLFEAKKWLRE-----GEWDQSERINAAKNCVLCIKYDPDDYENPCGNSNSAITEND 1110
QY      669 MNHACMLFVENYGOKIIKKNLCRN-FMLHLVSMHD---FNLISIMS-----IDKAVT 716
Db      1111 KNH-----LSEKQSYLSNWSMSYSYHPGFCHSESVYNKSEYLSKIDNSGIEBPVIK 1166
QY      717 KLEMQ-----OKLEKGESASPAN-----735
Db      1167 NIREKNTGFSEIMSPGREADTDPQSVNEDICVEKLTATNSCKNKNATKVAISDSNNFN 1226
QY      736 --EETTERQNTANGFSINSKE-----KALE--TDSVSGVSKQKQK 775
Db      1227 TIQKLSNDSNNSVPAYSIVNSKRVFAHQTKVTEGFTDNCNCSWTKQNTKSK 1277

RESULT 8
O59904
AC O59904 PRELIMINARY; PRT; 1314 AA.
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Silent information regulator 4.
GN SIR4.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198828; PubMed=9539421;
RA Astrom S.U., Rine J.;
RT "Theme and variation among silencing proteins in Saccharomyces
RE cerevisiae and Kluyveromyces lactis.";
RL Genetics 148:1021-1029 (1998).
DR EMBL; AF035007; AAC39438.1; -.
DR PIR; T09481; T09481.
SQ SEQUENCE 1314 AA; 146735 MW; 3C78CBB63319F8A6 CRC64;

Query Match 4.0%; Score 163; DB 3; Length 1314;
Best local similarity 19.9%; Pred. No. 0.008;
Matches 169; Conservative 140; Mismatches 338; Indels 204; Gaps 39;
QY 43 RVLEKQEL--QOPTYVALSYINRFTMDARREOBSLXK---KIOPKLSLTLSVSVS--RG 95

```

```

Db      181 RVVEVLRRQOSS--SSSGRIHKHIESPFQRMKPVSKIVKAPKNVAFAFETSUNDIDL 240
QY      96 NVSTPPRHSSGLTPPVTPPTTPSSSPRSSTTEPTQIYRFLRTN-----LIAPIEL 148
Db      241 NLLTRARASSGG-----TPTSGSAAVFAAP-----KTNNKTDQSVSKIANBIQ 284
QY      149 HRTLYMHRNRTWIKKTFKVDDMLSKVEKMKGEQESHLSAHLQLTFTGFFHKNDKP 208
Db      285 NNKGRISKANS--PIQRTTTTTSNIPSEAKEL-----YNMLSSSLKRRNVLFDKKGTIN 337
QY      209 SPNSNEQNSVTLEVLVVKVCHKKGVSCPIRQVPTKQKVPLIPDLNQTKPGFPFSLA 268
Db      338 SENNDKTGASVKVESHVVEILGSKEDHTHTSFVLPSPMQATIADLAKTITSLVKENSAQIE 397
QY      269 VSSNFEFEPNSHMVKSYSLLFRVTPGRREPFNGMGTENENIDVNEEL-----PARR 321
Db      398 VDRNLIGPERS-----RKNKEFQDL---ETRTSSPAKESLASPSADTPERK 441
QY      322 KENREDGEKTFVAQMTVPDKNRRLLQLLDGEY---VAMQMEECPI8-----365
Db      442 VPQLDLTSPSPDSTPGVTEKMDLKEVPSPVPSSAGAVQSSSQISHVITQAGADSG 501
QY      366 -----KKRATWETILDGKLPFPFETFSQGTLOFTLRWGETNDKSTAPIAKPLATR 417
Db      502 SHHGQIVESSELNIPTEPEKQNP-----QAASASGTFHHVDE--KQSTBETAKLLALS 553
QY      418 NESLHQNCKPGS-----VKPTQTIIVAKESLTLTDLQTRKEKTPNENRQKLRIFYQFLYN 472
Db      554 DTSS--EYSSDSSSGSHYVGDFFQLNTWKARWGPVAKWAF---NKSFPKDLMKYLIA 608
QY      473 NNTROQTEARDDLHCPWCTLNCRLKLYSL-LKHLKCHSRFIF-----NYVYHPKGA- 522
Db      609 EHT--PLAARSDVSCNDICNIRNYESVEVHLOKLPRLAYLELESPPSNNFISKPLRAN 666
QY      523 -RID-VSINECDGSGYAGNPDIHQPGFAPFASRNGPVK-----RTPIIT 563
Db      667 NRVEKLRKDELAYNGS-----KQKRTGAPDSESGKKPLIERQIEDGNVSEKTPQEPTT 721
QY      564 HLIVCRPKRTKASMEFELESDGEVEQRTYSNGHNRLYFHSDT---CLP-----610
Db      722 SI-----PGETDVNEKVLPIED--IEMLPFTFISSNRLEPDANTLSAHLFPSSVIGVKQ 775
QY      611 -----LRPQMEVDS---DEKDEWLRREKTTIQIEFSDVNEGEKVM 651
Db      776 PSLTTVKNKFPASFENGLOPDNAQASTEIETNEGNSKKGKDELHFDPERCEDVTQFRKEYL 835
QY      652 KLNHLVWKHGFADNQNHACMLFVENYGOKIIKKNLCRNFMHLVSMHDFNLISIMSI 711
Db      836 AL-RASISKDDTCSSSLVN-----EDLNKTEVIQNLTSII--QSEWKIASLVGI 882
QY      712 DKAVTKLREMOKLEK--GESASPANEIITBEQN---GTANGFSEINSKEALETDSVS 765
Db      883 NH---QLREKLEBELEKIQAQLVQELDTVLLESKGSFSAGTAK-AEVNNKDK--TESIN 935
QY      766 GYSKQSKQKQL 776
Db      936 --NEQAREKSL 944

RESULT 9
Q9U4X0
ID Q9U4X0 PRELIMINARY; PRT; 2647 AA.
AC Q9U4X0;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative erythrocyte binding protein EBL-1 (Fragment).
GN EBL-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]

```



```

QY 461 --OKLRIFQFLYNNTRQ---QTEARD-----DLHCPWCTLNCRLKLSL-LK 502
Db 1195 EVSKLOSEIQ-----NTQALKKLETRVVDLSKYKATKSDELTQISDLN-EKANLNK 1238
QY 503 HLKCHSRFIRVYVHPKARIDVINECYGSGVAGNPQDIHQPGFAPSRNGVPVKTPPI 562
Db 1239 YEEVCEE-----VLHAK--KKELSAKDEKELLHFSIQEIKDQ-----QBRCDKSLTTI 1285
QY 563 THILVCRPKRKASMSFLESED-----GEVQQRQYSSGHNLYFHSDTCLPLRPOE 615
Db 1286 TEL-----QRRQESAKQ--IEAKONKITELNDVRLKQALNGLSQLYVGSQ---PSRQS 1338
QY 616 MEVDSSEKDPWELREKTIQIEEFSDVNEGEKVMKLNJHVM 659
Db 1339 QLIDS-----LQQQVRSLLQQQLADADRQHQEVIAIYRTHLL 1374

RESULT 11
Q8MKI9 PRELIMINARY; PRT; 3471 AA.
AC Q8MKI9
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Brc2a protein.
GN BRC2a.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21570986; PubMed=11714026;
RA Ochiai K., Morimatsu M., Tomizawa N., Syuto B.;
RT "Cloning and Sequencing Full Length of Canine Brc2 and Rad51 cDNA.";
RL J. Vet. Med. Sci. 63:1103-1108 (2001).
DR EMBL; AB043895; BAB91245.2;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR Pfam; PF00634; BRC2_1;
DR PROSITE; PS00152; ATPASE ALPHA BETA; 1.
DR PROSITE; PS00138; BRC2_REPEAT; 7.
SQ SEQUENCE 3471 AA; 385085 MW; 4AB77908A8EF6561 CRC64;

Query Match 3.9%; Score 158; DB 6; Length 3471;
Best Local Similarity 18.2%; Pred. No. 0.069;
Matches 177; Conservative 146; Mismatches 364; Indels 284; Gaps 44;

QY 23 PPTLADLIEHDNHDIDTPVLEKQELQOQTYVA--LSYINRFMTDAARQESLKKKIQ 81
Db 1023 FKTASKEIKUSENNIRKSKMLFKDIEHYTNLACUEIVNTSLESQKPKSKS--HALD 1080
QY 82 PKLSLTLSSSVRGNVSTPPRHSSGSLPPVTPPTITSSSPRSSTPEPTQI----- 133
Db 1081 PQSINIISGFVQ--NSTVSDSESGHTAPPTLSLKQPDFSNRLNTPSQKAITELSTILE 1138
QY 134 -----YRLRNLNLIAPFLHRLTYMHSRNT--NKRKTFKVDY-----LSKV 178
Db 1139 ESGSQFETQPRK---PSGHIQRNPFEMPENQLNLTNSKWKXDDDLHLTTNAPSISQV 1195
QY 179 EMKK-----GEQSHSLSA-----HLQLTFTGFFH--KNDKSPNSE----- 213
Db 1196 DSKKEGILGKQKACLSRLSCNRSASGYSTDKNEVEFRGFYARGTKLVNVSALQKA 1255
QY 214 -----NEQNSV-----TLEVLVVKCHKRKNVSCPIRQVPTGKKQV 250
Db 1256 KKLFDLENINEETSVEVDRSFSSSKYNDVSVMIQIEDCNK----- 1297
QY 251 PLIPDLNOKPFPPLAVSNNEPSPSHVMKYSILFR-----VTRPGRREFNG 301
Db 1298 -----NLNEK---NNKCRILLQNNIEMTMDIFVEYTESYRNTENEGNQCTDAGNTCS 1350

```

```

QY 302 MING---ETNENIDVNEE---LPARKRNREDGKTFVAQMTVFQDKNRRLQLLDGEYEV- 354
Db 1351 ESDGSDSSKNDIVYIHEENGILPCIDQHNID-----LKLFSQFMKEGNTQIKEGLSDLT 1404
QY 355 ---AVQMECEPISKRRATWETILDGKRLPPPTFP--SQGPTLQTLRWETNDKSTAP 409
Db 1405 CLEVKAETSTWMTSNKQQLTANTQNIKDPDTFYLSFQTASRNINRVSRSLNKAESL 1464
QY 410 IAKPLATR---SESLHGENKPG--SVKPTQTIAVKESLTTDLOTRKEDTPNENRQKLR 464
Db 1465 LKQKTEBELNPNFSLNSSELLPGIDIKKD---ISNHEVIENTERKOKITKESLIGTE 1521
QY 465 IFYQFLYNNTRTQTEARDDHLCPWCTLNCRLKLSLKLHLKLCHSRF-----IFNY 515
Db 1522 NILLILQORPESKIKKIKESAVLGFHTASGKKIETKESLDKVKNLFEKEQDNSEITNF 1581
QY 516 VHPKGARIDVSNICYDG--SYAGNPQDIHQPGFAPSRNGVPVKETPTH-ILVCRP--- 570
Db 1582 SH--RGAKMSKDRBECKDQRELAOCGTTTITTPYEETHSSLEKKLVSNELIALRPLL 1639
QY 571 ----KRT-----KASMSBFESEDEG---VEQQRYS-----SGHNR 600
Db 1640 SDNLYKQTEENLKSIDHASQKVDVHENTEKETAKKPTMTNQSTYSAIENSPLTFYTGHR 1699
QY 601 LYFHSDTCLPLRPQMEVDSDEKQPEWLRKTIQIEEFSD-----VNEGEKVMKLV--- 654
Db 1700 KISVSEASL-----FEAK---WLRE-----GSWDDQSRINAANKVCLUEPYD 1740
QY 655 -----NLHVMKHGFIADNOMHACMLFVENYQGKIIKKNLCRN-FMLHLVSMHD---FN 704
Db 1741 DIVENPSCGNSNSAITENDKNH-----LSEKQGSTYLSNSTMSNSYSYHPGFCHSSEYN 1796
QY 705 LISIMS-----IDKAVTKLRMQ-----OKEKGE 729
Db 1797 KSEYLSRSKIDNSGIEPVIKTRERKNIGFSEINSFGREADTDPQSVNEDICVEKLATNS 1856
QY 730 SASPAN-----EETEEQNGTANGFSEINSKE-----KALE--TDSV 764
Db 1857 SKCKNRTAKVAISDNNFNFTIQKLNSDNNNSVPAYSVNKRVFVAHQTKVTEGFTDNC 1916
QY 765 SGVSQSKQKQK 775
Db 1917 SMVTKQNTKSK 1927

RESULT 12
Q8I525 PRELIMINARY; PRT; 5767 AA.
AC Q8I525;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein.
GN PFL1930W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=222555705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Paine A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pitera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum.";

```





```

Db      650 LKASNAVEKIILKLTQEDHYKLGIIIRFKUKYEHSHIESFDQIKLDPKHKKALHNKGI 709
QY      707 SIMSDKAVTLKREMQ--KLEKGSASPAHEEITEBQNGT-----ANGPSEINSKE 756
Db      710 ALMLNKNKAIESFEKAQIDKNYSTAYYKGIKAEKNGDMQQAFAESFKAYNLDKKPN 769
QY      757 KALETDSVS--GVSKQSKK 773
Db      770 YALKAGIVSNLGNFKQSEK 789

RESULT 14
Q8IKP5 PRELIMINARY; PRT; 1163 AA.
AC Q8IKP5;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF14_0859.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7; PubMed=12368864;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Shellen J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian L.G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511 (2002).
DR EMBL: AE014825; AAN37172.1; -
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00056; zf_C2H2; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 1163 AA; 139588 MW; 6B37EB4621D0DAB1 CRC64;

Query Match
Best Local Similarity 3.8%; Score 157; DB 5; Length 1163;
Matches 170; Conservative 127; Mismatches 303; Indels 308; Gaps 47;

QY 19 CGLHFEPTLADLEHIEDNHIDTPRVLEKQBLQOPT-----YVALSVIN----- 62
Db 271 CGLRPMTSMKLNHLE-NHYKKSQFYLNNSSRSFTKSKKFLYLDHINLPIEVFVCKNY 329
QY 63 -----RFM-----TDAAREQESL-----KKKIQPK 83
Db 330 SIFEDFYNVTKNMAFNQGSHEIEQILDNDNDNSYHNDDEKIYIEGEGEKKKDESK 389
QY 84 LSLTLSSSVSRGNVSTPRHSSGSLTPVTPPTPSSFSRSTTEPTQIYRFLRTRNLI 143
Db 390 NNDLSYSL-HTKITHFNKDN-----QESFLNN-----NTRKIQ 424
QY 144 APIFHLRLTYMHSRNSNTK-----RKTFKVDMLSKVKMKGEBSHLSHLQITF 198
Db 425 EKYNNYNNLSNNNNNNNDVANLNFNFKKTYNNID----- 459
QY 199 TGFHFNKDKPSNNEQNSVTLEVLVKVCHKKRQKQVSCPIQVPTGKQVPLIPDLNQ 258
Db 460 -NFFH-----NKETIHMLNPNMAVDK--NKKKNI-----LND 490
QY 259 TKPGNFPLAVSSNEFEPSNHHMVKSY-SLLPRVTRGRREFNGINGET-NENI----- 311

```

```

Db      491 REKG---SNEINFFDIYGNENTYDVLYSNYYVTEDNSVLIN-YINGPTMYRNIMKCLE 546
QY      312 --DV-NEELPA---RKRNRDEGEKTFVAQMTVFQKNRLQLLDGEYEVAMQMBECCIS 365
Db      547 IKDIYNYKFPSPWKESIN-----NFFIKRIEIDK-----LIQ 583
QY      366 KKEATWETILDGKRLPPFPETFS-QGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESL-- 422
Db      584 KKTNIFTNLSGR--TPSNTPFLNSYNIQDDLNNNNNNNNLKGKKNKIDIVINKNNLWN 641
QY      423 ---HQ-ENKPGSVKPTQTIAVKESLTTDLQ-----RKEKOTPN-----EN 459
Db      642 ICTTTFYNSKGLCPINVFLSYKENTLDHQTHSDKNDHDDKESNNYPMQNSICTQES 701
QY      460 RQKLRIFYQFLYNNNTROQTEARDDHLCPCWCTLNCRLKY-SLLKHLK-----LCHSRFI 512
Db      702 VENNHLVLSFL--NIFNQKYPACTDIFKNILLFHRKHYFMIINELKNONIDFNEULHSR-L 758
QY      513 FNYVHPKGARIDV-----SINECYDGSYAGNPQDIHRQPCGAFSRNPVKRTPTIHL 566
Db      759 FQYIYNNYSKTISIFNNINYKIDTCF-----LCKENFSFEYSVEYNDFFYTNVI 807
QY      567 VCRPKTKASMSEFLESEGEVEQORTYSSG-----HRLYFHSHTCLPL 611
Db      808 CVDLK-----NVYENDDDVEDTEEDANKNYDIDMSIKRIDHMCDEVYNNNSYDIDMKCLYE 863
QY      612 RPQEM-EVDSDEKDPWMLREXTITQIBEF-----SDVNE--GEKVMKLNWLMVKH- 661
Db      864 KHKELNELLYENKE---MSDDAITQVELLNNKMYNNEINECSNHNKKNVINTYDNI 920
QY      662 -----GFIADQNMHACMLFVNYGQKIKNKCRNFMHLVSK-----H 701
Db      921 NINNIFTDIKKIIDDN-----CFL-----GNTIIESNMDCNFHNNIKNKKKKDHIMNH 970
QY      702 DFNLSIMSIDKAVTKLRMOCKLEKSGESASPAHEEITEBQNGTANGSEINSKEKALET 761
Db      971 YNNLFKNLCIPYDTL---DILHRIKKG-----DITGKNTMTNNTNSNTNDEGYTN 1019
QY      762 DSVSGVSK 769
Db      1020 EEDNFKL 1027

RESULT 15
Q9W4G6 PRELIMINARY; PRT; 1893 AA.
AC Q9W4G6;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE CG2861 protein.
GN CG2861.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley; PubMed=10731132;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```



RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hostin D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Reinigen K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage I., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frise E., Galie R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003432; AAF45987.2; -  
DR FlyBase; FBgn0029728; CG2861.  
SQ SEQUENCE 1893 AA; 211119 MW; 3805732AD01E2098 CRC64;  
Query Match 3.8%; Score 157; DB 5; Length 1893;  
Best Local Similarity 19.3%; Pred. No. 0.036;  
Matches 164; Conservative 118; Mismatches 310; Indels 256; Gaps 40;  
QY 6 AASFFNTCTCFGCGGLHFTLADLIEHIEDNHDIDTPRVLEKQE-----LQOPTYVA 57  
DB 1145 AASRSSNGGTIGFCNLKAFRRDRPMISNAENTGF-TNPSELNKKCATKYCTPMERNSSGQ 1203

Search completed: August 25, 2004, 17:57:30  
Job time : 141 secs

QY 58 LSYINRFTWDAARQESLKKKIQPKLSLTSSSVSRGNVSTPPRHSSGSLTPVTPPIT 117  
DB 1204 LPIONRFILENSKEEG--KMWVEPVN-----PKHK-----PSRIPIR 1239  
QY 118 PSSFSRSTTEPTQIYRFLRTNLAPIFLHRTLTYMSHRNSTNI-----KCKT 168  
DB 1240 TKTPIALNLMKEPEDNAR--KPNKILAPFNSADVNPCSDTDEPNFKCESLPALAKND 1297  
QY 169 FKV--DDMLSKVKEMKGEQESHLSAHLQLTFTGFPHKNDKPS--PNSENEQNSVTLVLL 225  
DB 1298 FKLKNDKKSWSHYQ--EQESQELKEN-----SFFYAKFNLPINPKKSKANAEERC 1351  
QY 226 KVCKHKRKO-----VSCPIRQVPTGKKOVPL 252  
DB 1352 TKTFEDYKKAQSQRYTKAINKRQNKAMLESALGNAITDEPTVNCMLAONPDQVEMPR 1411  
QY 253 IPDLNQ--TKPGNPPLSLAVSNPEPSNSHMVK-----SYSLFLFRVTRPRRRFNGMNG 305  
DB 1412 NPDIVEITVPEKFAVPSAS--IKPKSSKKRKLMSASNSKLKF-----KIFSEVVK 1464  
QY 306 ETNENIDV--NEELPARRKRNRED-----GEKTFVAQMTVDKNNRLQLLDGEYEVAMQE 358  
DB 1465 DENEEEDVIPDNLIIPVKRKQNISELLDMQFSTSKAAATITSOFKIAQLKRMPSKLNEK 1524  
QY 359 MEECPISKKRATWETILDGKSL-----PPFETFSQGTTLQFTLRWTGET 402  
DB 1525 MKS--YAKKSGNDNDKDTKNTGMHMYKIPKLSANKPQNTSTTQSTFGFS--RLFQMY 1581  
QY 403 NDKSTAPIAKPLATR-----NSESLEHKNKPGSVKPTQTIIVKESLTTDLQTRKEKDTP 456  
DB 1582 QEPNFQNAQHPFSTRGYPYQRAYPLHRPTSQVHMNQPLLELHHSVMYDAEPSFESVFP 1641  
QY 457 NENRQ-----KLRFYQF-----LYNNNTFQOQTEARDLHCPWCTLNCR 495  
DB 1642 APYQPNINPMSMHHSYIPCEPSPSPYPMGGIPFQNNNT----- 1691  
QY 496 KLYSLXKLKLCHSRFIFNVYH---PKGARIDVSIINECYDGSYAGN--PQDIHRQPGFAF 551  
DB 1682 -LYPPMEN-DLCYPRGRNLNYPVRLPPMNSQ-----GTYPNYPQNF--LPNYPP 1727  
QY 552 SRNGPVRKTPITHLVCRPKRTKASMSEFLESDGEVEQOQRTYSSGHNRLYFHSDDTCLPL 611  
DB 1728 SRQFLSLMQRLADI-----NKQIASGSIESVGFI--QWYAGRYR-----TLKDL 1772  
QY 612 RPOBMEVDSEDE-----KDPWLREK-----TITQIEEFS---DVNEG 646  
DB 1773 RERTML--SKDSLAKIPKCMCKIKRPAWRISKFYLYAYWCEISCIPOSSBFGKPLENAA 1830  
QY 647 EKEYMKLNLHVMKHGFIADNQ-MNHACMLFVENYQOKIIXKNL-----CRNF 693  
DB 1831 YRKVISSY-----AENKFLRMSLSMEKTGFGKSKLKVLMIGRKRSTKWRLLNY 1880  
QY 694 MLHLVSMH 701  
DB 1881 RKPLIHLH 1888



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:46:43 ; Search time 45 seconds  
(without alignments)  
1659.769 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 4078  
Sequence: 1 MTGIAAASFNTCRFGGOG.....KALETDSVSGVSKQSKQKL 776

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*

1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176.5	4.3	1142	2 S59359	GIN4 protein - yea
2	166	4.1	1507	2 T42631	breast cancer tumo
3	163	4.0	1314	2 T09481	mating type silenc
4	155.5	3.8	1755	2 S57045	TyB protein - yea
5	154	3.8	1271	2 A45555	glutamate rich pro
6	152	3.7	1119	2 B70126	surface-located me
7	152	3.7	1749	2 S69972	TyB protein - yea
8	152	3.7	1957	2 T38077	hypothetical coile
9	150	3.7	912	2 B44450	ubiquitin-specific
10	149	3.7	1145	2 T18235	transcription acti
11	148	3.6	863	2 T51002	hypothetical prote
12	147.5	3.6	1368	2 T18371	probable glutamate
13	147.5	3.6	1755	2 S63980	TyB protein - yea
14	147.5	3.6	1790	2 S67593	potassium transpor
15	146.5	3.6	1235	1 PWSYH	tanabin - African
16	145	3.6	1744	2 JH0720	BN11 protein - yea
17	145	3.6	1953	2 S63244	hypothetical prote
18	143.5	3.5	698	2 S45027	probable major sur
19	143	3.5	1640	2 A24594	TyB protein - yea
20	142	3.5	1278	2 B24671	interaptin - slime
21	142	3.5	1738	2 T14466	hypothetical prote
22	141.5	3.5	1263	2 T15496	TyB protein - yea
23	141.5	3.5	1328	2 S52481	TyB protein YER138
24	141.5	3.5	1755	2 S50641	TyB protein - yea
25	141.5	3.5	1755	2 S63975	hypothetical prote
26	141	3.5	1827	2 T18270	myosin-like protei
27	141	3.5	1875	2 S38173	hypothetical prote
28	140.5	3.4	4688	2 F82885	major merozoite su
29	140	3.4	1631	1 SAZQK1	

30 139.5 3.4 1085 2 S62516 hypothetical coile  
31 139.5 3.4 1755 2 S69949 TyB protein - yea  
32 139.5 3.4 2500 1 MWHE2 HIV-SP2 enhancer-b  
33 138.5 3.4 582 2 E84771 fertilization-inde  
34 138.5 3.4 1755 2 S50663 TyB protein - yea  
35 138.5 3.4 1793 2 S52601 TyB protein - yea  
36 137.5 3.4 1196 2 S42098 TyB protein - yea  
37 137.5 3.4 1755 2 S69839 TyB protein - yea  
38 137.5 3.4 1755 2 S69960 TyB protein - yea  
39 137.5 3.4 1755 2 S69979 TyB protein - yea  
40 137.5 3.4 1755 2 S69955 TyB protein - yea  
41 137.5 3.4 1756 2 S69983 TyB protein - yea  
42 136.5 3.4 2954 2 T14156 kinesin-related pr  
43 136.5 3.3 791 2 T20815 hypothetical prote  
44 136.5 3.3 1755 2 S70238 TyB protein - yea  
45 135.5 3.3 1295 2 T24587 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S59359  
GIN4 protein - yeast (Saccharomyces cerevisiae)  
N/Alternate names: protein YDR507c  
C/Species: Saccharomyces cerevisiae  
C/Date: 30-Nov-1995 #sequence\_revision 01-Mar-1996 #text\_change 24-Sep-1999  
C/Accession: S59359; S69565  
R/Longline, M.S.; Pringle, J.R.  
submitted to the EMBL Data Library, August 1995  
A/Reference number: S59359  
A/Accession: S59359  
A/Molecule type: DNA  
A/Residues: 1-1142 <LON>  
A/Cross-references: EMBL:U33140; NID:g992650; PIDN:AAA75513.1; PID:g992651  
R/Dietrich, F.S.  
submitted to the EMBL Data Library, August 1995  
A/Description: The sequence of S. cerevisiae cosmids 8166, 9787, 9717, and lambda 3073.  
A/Reference number: S69553  
A/Accession: S69565  
A/Molecule type: DNA  
A/Residues: 1-1142 <DIE>  
A/Cross-references: EMBL:U33057; NID:g927764; PIDN:AAB64949.1; PID:g927777; MIPS:YDR507c  
C/Genetics:  
A/Gene: SGD:GIN4  
A/Cross-references: SGD:S0002915; MIPS:YDR507c  
A/Map position: 4R  
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
C/Keywords: ATP  
F/17-289/Domain: protein kinase homology <KIN>  
F/25-33/Region: protein kinase ATP-binding motif

Query Match 4.3%; Score 176.5; DB 2; Length 1142;  
Best Local Similarity 20.8%; Pred. No. 0.0037;  
Matches 152; Conservative 102; Mismatches 264; Indels 213; Gaps 35;  
QY 41 DPRVLEKQEIQQP-----TVVALSYINRFMTDARREQSLKKKIQPKLSLTUSS-SVS 93  
DB 337 DPEGI-KEKLREPAGNAEKLTYALLY--RPFKCDT---OKELIKQQQVKRQSISSVSVS 389  
QY 94 RG-NVSTPP-----RHSSGSLTPVTPPI-----TPSSSFRSSTPTPTQIVRFLRTR 140  
DB 390 PSKVVSTTPQRRNRRESLIVTSRKPPIFNKFTASSASSSNLTTPGSSKRISKNFSSK 449  
QY 141 NLIAPIFLHRTLTVMYSHRNRTNI--RKTKFKVDDMLSKVERKMGQES----- 187  
DB 450 KKLSTIVNQSSPTPAS-RNKRASVINVEKQKQASIFSTTKNKRSSRSIKRMSLIPSMK 508  
QY 188 -HSLSAHLQLTFTGFFHKNDKPSNSNEQNSVLEVLVAV-----CHKKRKDVSC 238  
DB 509 RESVTTKLMSTYAKLAEDDDWEYIERTKRTSSNFATLIDEIFEYKYEQIRKEKELE 568  
QY 239 PIQVPTG-----KKQVPL-IPDLNQ----- 258

```

Db      569 KVRKAKRELEERRRRKQEKERARKLLEKEDLKRKQBELKKQIETIDISLEQELSKHKE 628
QY      259 -TKPNNFSLAVSNEFPNSHMVKSYSLLFR-----VTRPGRRFNGMNGINGETNENI 311
Db      629 EKLGNIRISAPMEENEKKNINHLVDINILRRRNFSLOTRPVSRDLPDGMFMSPTTEEV 688
QY      312 DVNEELPARKKRENGDEGKTFVAQM-----TVFDKRRRLQLLDGEYE--VANQEM-EE 361
Db      689 SPVE--PKRTENERUTTKKILLETIRRSKFLGSSFNIDKELKMKMEYPSIIAQRISEE 746
QY      362 CPISKRRATWETIL--DGKRLPPPE--TFSGQPTLOFTLRWTGETNDKSTAPIAKPLAT 416
Db      747 RVVSDNDGYESLILPKDNGVSQLKSTATTAPVSDGRLKISE-----IRVQFT 798
QY      417 RNSBSLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTNPNRQKRLIFYQFLYNNTR 476
Db      799 RKSRSFSESNKLSV--LSMYSYTKESFTNLVDILKNGLDVNNQSQRI-----845
QY      477 QOTEARDLHCPWCTLNCRKLYSLKLHLKCHSRFIFNYVYHPKRGARIDVINECYDGSY 536
Db      846 PTPRSADD-----SEELFE-----TVNE--EABY 867
QY      537 AGNPQDIHR-----OPGFA--FSRNGPVKRTPIITHILVCRPKRTKAMSE 579
Db      868 TGNSSNDERLYDVGDSTIKDSALKLNFAFRNGSNEAKQTDNLHLPLILPLINGDNLK 927
QY      580 FLESDGEVEQRTYS---SGHNLVPHSDCLPLRQPEMEVDSDEKPEW-----628
Db      928 -QNSGEQDAHPKIKSMIPESGSS---HTE-----KEENEKEKKEPQKHQBEDQ 976
QY      629 -LREKTIQIE 638
Db      977 EKREKVDDME 987

RESULT 2
T42631
breast cancer tumor suppressor BRCA2 - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 24-Nov-2003
C:Accession: T42631
R:Bignelli, G.R.; Micklem, G.; Stratton, M.R.; Ashworth, A.; Wooster, R.
submitted to the EMBL Data Library, July 1996
A:Description: BRCA2 contains a novel repeat that is conserved in other mammian species
A:Reference number: Z22184
A:Accession: T42631
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1507 <BIG>
A:Cross-references: EMBL:Z75664; NID:e996328; PID:e251256; PIDN:CAA99994.1
C:Genetics:
C:Superfamily: breast/ovarian cancer tumor suppressor protein, BRCA2 type
C:Keywords: tumor suppressor

Query Match      4.1%; Score 166; DB 2; Length 1507;
Best Local Similarity 18.9%; Pred. No. 0.025;
Matches 180; Conservative 143; Mismatches 384; Indels 244; Gaps 43;

QY      23 FPTLADLIEHDNHDIDPRVLEKQELQOQTYVA--LSYINRFMTDAARRQESLKKKIQ 81
Db      373 FKTASNKIEKLSNNIRKSKMLFKDIEBHYPNLACLEIVTSSLESQKPKSKS--HALD 430
QY      82 PKLSITLSSSVRGNVSTPPRHSSGSLTPVTPPTPSSSPRSSTPTPEPTOI-----133
Db      431 POSINIIISGFVQ--NSTVYSDSESHTAPPTLSLKQDFSDSNRNLTSPQKABITELSTILE 488
QY      134 -----YRLRTNLIAPFLHRTLYMGHRSRT--NIKRTTFKVDMM-----LSKV 178
Db      489 ESGSFETQTPFK---PSHIQKNPFEMPEQLILNLSKEWKDDDLHLTNAPSISQV 545
QY      179 EKMK-----GEQSHSLISA-----HLQITFTGFFH-KNDKPSFNSE-----213

```

```

Db      546 DSKKSEGIIGKKQKQACLSRTSCNRSASGYSTDKNVEFRGFYSARGTKLVNYSSEALQKA 605
QY      214 -----NQNSVTLEVLVVKVCHKKRDVSCPIRQVPTGKQKQVPLIPDLNQTQPN 263
Db      606 KKLFSLDLENIBETSVEVD-----RSFSSSKYNDVSMIQIEDCNDKNINEPNN 654
QY      264 FPSLAYSSNEFPNSHMVKSYSLLFR-----VTRPGRRFNGMNG--ETNENI 311
Db      655 KRL--ILQNNIEWTDDIFVEETESYRRNTENEGNOCTDAGNTCNSRSDGSDSKNTIV 713
QY      312 DVNEE--LPARRKRNREDGEKTFVAQMTVFQDNRLQLLDGEYEV-----ANQEMEBQPI 364
Db      714 YIHEENGSLPCIDQHNID-----LKLFSQFMEKGTQIKEGLSLDLCLEVMKABETSHV 767
QY      365 SKKRAWTETILDGKRLPPPETF--SQGPTLOFTLRWTGETNDKSTAPIAKPLATN----418
Db      768 TWSNKOQLTANTGONIKDFTFYLSPQASRKNIIVRSLSLNAKSLNKKQWTEBELANF 827
QY      419 SSSLHQENKPG-SVKPTQTIAVKESLTTDLQTRKEKDTNPNRQKRLIFYQFLYNNTRQ 477
Db      828 SDSLSNELLPGLDIKKTQ---ISNHEVIENTERKDKITKESDLIGTENILLILQORPESK 884
QY      478 QTEARDDLHCPWCTLNCRKLYSLKLHLKCHSRF-----IFNYVYHPKRGARIDVSI 528
Db      885 IKKIKESAVLGFHTASGKKIETIKESLQKVLFEKEQDNSEITNFSH--RGAKMSKDR 942
QY      529 NECYDQ-SVAGNPQDIHQPGFAFRNGPVKRTPIITH-ILVCRP-----KRT-----573
Db      943 ECKQORELACGTTEITTTPEYEETHSLEKKKLYSNEIALEPRLSLDNLKQYENLKI 1002
QY      574 -----KASMSFLESDGE---VEQRTYSNGHNR-LYFHSDTCLPLRQPEMEVDSDE 622
Db      1003 SDHASQKVDVHENTKETAKKPTMTVNTQSTYSAIENSPLTFTQDT-----EKEFSVSEA 1056
QY      623 E--KQPEWLREKTIQIEFSD-----VNEGEKEVMKLM-----NLHVMKHGFIADNQ 668
Db      1057 SLFEAKKWLRE-----GEWDDQSERINNAKVNCLKEYPDDIVENPSCGNSNSAITEND 1110
QY      669 MNHACMLFVENYQKIKKLNCRN-FMLHLVSMHD---FNLISIMS-----IDKAVT 716
Db      1111 KKH---LSEKQGSTVLSNSTMSNGSVHPGCHSSEVNVKSEVLSRSKIDNSGIEPVIK 1166
QY      717 KUREMQ-----QLEKGESASAPN-----735
Db      1167 NIREKNIQFSGIMSFGREADTDPQSVNEDICVEKLATNSSCKNKNKTAIKVAISDNNFN 1226
QY      736 --EETEEQNGTANGFSEINSKE-----KALE--TDSVSGVSKQSKQK 775
Db      1227 TIQKLSNDSNNSVPAYSTVNSKRVFAHQTKVTEGTDNCNMTVKQNTKSK 1277

RESULT 3
T09481
mating type silencing regulator SIR4 - yeast (Kluyveromyces marxianus var. lactis)
N:Alternate names: silent information regulator 4
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T09481
R:Astrom, S.U.; Rine, J.
Genetics 148, 1021-1029, 1998
A:Title: Theme and variation among silencing proteins in Saccharomyces cerevisiae and Kl.
A:Reference number: Z16686; MUID:98198828; PMID:9539421
A:Accession: T09481
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1314 <AST>
A:Cross-references: EMBL:AF035007; NID:g3002787; PID:g3002789
A:Note: the source is designated as Kluyveromyces lactis
C:Genetics:
A:Gene: SIR4
C:Function:
A:Description: involved in both silencing and telomere length maintenance

```



```
QY 610 PLRPOEMVDSDEKDPWLRKTIQTIEEFSD-----VNEGEKE 649
      : : : : : : : : : : : : : : : : : : : : : :
Db 1423 CLFVDDMLFSKDLN-----ANKKIITLKKQYDTKIINLGSD 1461

RESULT 5
A4555
glutamate rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C:Accession: A4555, 827831
R:Porre, M.B.; Dziegiel, M.; Hogn, B.; Petersen, E.; Rieneck, K.; Riley, J.F.;
J.
Mol. Biochem. Parasitol. 49, 119-131, 1991
A:Title: Primary structure and localization of a conserved immunogenic Plasmodium falciparum
rate life cycle.
A:Reference number: A45555; MUID:92131041; PMID:1775153
A:Accession: A4555
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1271 <BOR>
A:Cross-references: EMBL:M59706; NID:g160311; PID:g160312
A:Note: sequence extracted from NCBI backbone (NCBI:77801, NCBI:77802)

Query Match 3.8%; Score 154; DB 2; Length 1271;
Best Local Similarity 18.0%; Pred. No. 0.11;
Matches 150; Conservative 123; Mismatches 270; Indels 290; Gaps 40;

QY 31 EHIENHIDTDPRVLE-----KQELQOPTYVALSYINRFTMDAARREQESLKKIKQ 81
      : : : : : : : : : : : : : : : : : : : : : :
Db 258 EHLKDEKIENELVHENLIPNDIEQILNQ-----EQET---NIQ 296

QY 82 PKLSLTSSVSRGNVSTPPRHSSGLTPPVTPITPSSFRSSTPTPTPTQIYRFLTRN 141
      : : : : : : : : : : : : : : : : : : : : : :
Db 297 EOLYNEKONVEKQNSQIP-----SUDLKEPTNEDILP----- 330

QY 142 LIAPFLHRTLYMHRNRTN-IRKTFKVDMLSKVEKMG--EQBSHLSLAHLQLT 198
      : : : : : : : : : : : : : : : : : : : : : :
Db 331 -----HNPLENIQSESEINHVDHALPRENIIDKLDNKEHIDQSQHNINVLQENN 382

QY 199 TGFPHNDKPSNSEN-BQNSVTLEVLVVKVCHKKRDVSCPIRQVPTGKKQVPLIPDLN 257
      : : : : : : : : : : : : : : : : : : : : : :
Db 383 INNHOLEPQEKPNIESFPKNIDSIIL-----PENVETEBIIDVP 424

QY 258 QTKPGNFFSLAVSNFE-----PSNSH-MYKSYSLFLFRVTRPGRREFGNMGINGTEN 310
      : : : : : : : : : : : : : : : : : : : : : :
Db 425 SPKGNHETFEETSESEHEEAVSKNAHETVEHEETVQSENPEKADNDGNSQNGNE 484

QY 311 IDVNEELPARRKRNREDGEKTFVAQMTVDFNRRLLQDLGDEYE--VAMQEMEECPISKRR 368
      : : : : : : : : : : : : : : : : : : : : : :
Db 485 LNEENFVES-EKSEHEAAE-----NEESSLEEGHHEIVPEONNEESGESK-- 529

QY 369 ATWEIILD-----GKLPPEFTEFSQPTQLTIRWTGETN-----DKS-TAPI----- 410
      : : : : : : : : : : : : : : : : : : : : : :
Db 530 -----LVNDDEGGFEAHHEFNFSSEVS-----NSELNENEFVESKSVTEPAHEBEV 577

QY 411 ---AKPLATRNSESLHQNKPGSVKPTQTIAKESLTTDLQTRKEK-----TPNE 458
      : : : : : : : : : : : : : : : : : : : : : :
Db 578 SEESNPEPAENEESEIEAHQEEIYPEQN--DBESGESGLVDNBEQPEEPNHEEFPDQ 635

QY 459 NQOKURIFYQFYNNNTQQTEARDLHCPWCTLNCRKLYSLHLKLCRFRFTFNVTYH 518
      : : : : : : : : : : : : : : : : : : : : : :
Db 636 NDSEL-----SENELVESE-----KSVSEPAHEHVEIVSEKSVSEPAEH 673

QY 519 PKGARIDVINECYSGVAGNPQDTHROPGFAPS--RNGPVKRTPIITHI--LVCRPKRTK 574
      : : : : : : : : : : : : : : : : : : : : : :
Db 674 -----VEI-VSEKTSPEAHVESVSESNNEPEKKGVPSPFPFEEIKVDVQPKIVD 727

QY 575 ASMSE--FLESDDGVBOORTYSSGHNRLYFTHSTCTPLRPOEMEVD-----SEDBK-----D 625
      : : : : : : : : : : : : : : : : : : : : : :
Db 728 LQIIEPNFVDSQNPQE-----PVEPSFVKIEKVPSEENKHAUSD 767

QY 626 PEWLREKTIITQIEEFS-----VNEGEKEMKLNHLVHMKGFIADNQNHACML 675
```

```
      : : : : : : : : : : : : : : : : : : : : : :
Db 768 PEVKEKENVSEVVEEKQSQESVEEIPVNEDEFVHTEQ-----DLDH----- 812

QY 676 FVENYQKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOOKLEKGESASPAN 735
      : : : : : : : : : : : : : : : : : : : : : :
Db 813 -----KTVDFEIVEVEEIPSELHENEVAHFEI 839

QY 736 ERITE-----EQNGTANGFEINSKEKA-----LETDSVSGVSKQSKOKL 776
      : : : : : : : : : : : : : : : : : : : : : :
Db 840 VEIEEVFPENQN---NEFQEIENEDDKSAHQIEHVEVEEI--LPEDDKNEKV 887

RESULT 6
B70126
surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 11-Jan-2000
C:Accession: B70126
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Yugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98055943; PMID:9403685
A:Accession: B70126
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1119 <KLF>
A:Cross-references: GB:AE001131; GB:AE000783; NID:g2688098; PID:AAC66595.1; PID:g268810
A:Experimental source: strain B31
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hom
F;775-808/Domain: tetratricopeptide repeat homology <TT2>
F;742-774/Domain: tetratricopeptide repeat homology <TT3>
F;809-842/Domain: tetratricopeptide repeat homology <TT4>
F;843-876/Domain: tetratricopeptide repeat homology <TT5>
F;911-943/Domain: tetratricopeptide repeat homology <TT6>
F;944-977/Domain: tetratricopeptide repeat homology <TT7>
F;979-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>
F;1011-1044/Domain: tetratricopeptide repeat homology <TT8>

Query Match 3.7%; Score 152; DB 2; Length 1119;
Best Local Similarity 18.1%; Pred. No. 0.13;
Matches 147; Conservative 131; Mismatches 312; Indels 220; Gaps 33;

QY 74 ESLKKI-----QPKLSLTSSVSRGNVSTPPRHSSGLTPPVTPPI 116
      : : : : : : : : : : : : : : : : : : : : : :
Db 198 EKLKMLSKSSENILNDSQKIENDKQNTLUSKKNSENILKTFDNKYSNNNTSLK 257

QY 117 TPSSSRSTPTPTQIYRFLTRNLIAPIFLHRTLYMHSNRTNIRKRTFKVDDMLS 176
      : : : : : : : : : : : : : : : : : : : : : :
Db 258 KTSNSQKESLSP-----SQTIIIGKIY--RPVSYL-----IKKELYILDDIN 300

QY 177 KVEKMGGEBSLSLAHLQLTFTGFHKNDKSPSNEQNSVTLEVLVVKVCHKKRDV 236
      : : : : : : : : : : : : : : : : : : : : : :
Db 301 TCRVTILGKRLK-----ELIKKLSNPKQVKNELIENSKNEASNLTLTLI-----KKDI 350

QY 237 SCPIROVPTG--KKQVPLIPDLNQTKPGNFPSPILAVSNSEFEPSNHHVKSYSLLFRVTRP 294
      : : : : : : : : : : : : : : : : : : : : : :
Db 351 EPNLINIPDKPKKELFQL-DKEDKKPQVLEDLKSQVHSIKPIDLENTKSRQAIAK--- 405

QY 295 GRREFNGMNGETNENIDVNEELPARRKRNREDGEKTFVAQMTVDFKRRRLQLLGEVEV 354
      : : : : : : : : : : : : : : : : : : : : : :
Db 406 -----DLNEEL-----KNNPNDQ-----ASKLAQANKIQHLEDLKSQV 440

QY 355 AMQEMEECPISKRAETWETILDGKELPPF-----ETFSQPTQLTIRWTGETN 403
      : : : : : : : : : : : : : : : : : : : : : :
Db 441 --HSIKPIDLENTKSRQAIAKD---LNEFLKNNPNDQAASKTLAQANKIQHL-----ED 489

QY 404 DKSTAPIAKPLATRNSESLHQ-----ENKPGSVKPTQTIAVK-----ESLTDLQ 448
      : : : : : : : : : : : : : : : : : : : : : :
Db 490 LKSKVHSIKPIDLENTKSRQAIAKDINEFLKNNPNDQAASKTLAQANKIQHLEDLKSQV 549
```



```

Db 991 DTEHLKKNVSEVEVERNALLASNERLMDLKNNGENIASLQTEI-----EKGAENDDLQS 1046
QY 239 PFRQVPTGKQVPLIPD-----LNQTK--FQNFPSLAVSNPEFSPSHVKSYSYL 287
Db 1047 KLSVVSSEVENLLLSQTNKSLEKTKNQLYIEKNQVLDKDEKQNVLEBEETSKYK 1106
QY 288 LFRVTRPGRBFBNGINGETNENIDVNEELPARKKRNREDGE--KTFVAQMTVPDKRRRL 345
Db 1107 L-----GS--ENAIQKDELLALRKKSKQHDLCANFVDDLK--EKS DAL 1146
QY 346 QLLDGEYEVAMQEMECFISKRRATWETILDKRLPPETFSQFTLQTLWTGETNDK 405
Db 1147 EQLTNEKRELIVSLQ-----SNNNE 1168
QY 406 STAPIAKPLATNSE--SLHQENKPGSVKPTQTIIVKESLTTDLQTRKEKDTNENRQK 462
Db 1169 ALVEERSDLANLSDMKSLSDSDNVISIRSDLVVNDELDT---LKKDKDS----- 1218
QY 463 LRIFYQFLYNNNTROQTARDLHCPWCTLNCRKLYSLKHLKLCRGRFIPNYVYHPK-- 520
Db 1219 -----LSTQYSEVCQDRDD-----LLDSLKGCEBSF-NKYAVSLREL 1254
QY 521 --GARIDYSINECYDGSX--AGNPQDIHR-----QPGF----- 549
Db 1255 CYKSEIDVPVSEILDNFFVAGNFSELSRLTVLSLENYLDAFNQVFKMELNRLTYY 1314
QY 550 --AFGR-----NGPVKRTPIITHILVCRPK---RTKASMBEFLS-EDGEVEQ 590
Db 1315 DAFTKVVADLEKLOHEHDDMLIQGLDEKALKDSEKNFLRKEAEMTNIHSLGKEET 1374
QY 591 QRTYSNGHNLRYFSDTCLPAPQEMEVDSDEKDPWLRK--TIQIIEFSDVNEGK 648
Db 1375 KKETAELSRLEDNQALTNKLNQIDHLNQIRKEDVLKESLISLEP-SLSNQORQ 1433
QY 649 EVMKLNVLVVKHGFADNMNMHACMLFVENVGQKIIKKNL-CRNFMLHLVSMEDFLNIS 707
Db 1434 ESSLDAKNELEHMLDDTSRKNSSLMEXIESINSLDDKSPELASAVEKLGALQKLSHES 1493
QY 708 IMSIDKAVTKLREMOQKLEKGESE-SPANEBITEQNGTANGPSEINSKEKAL 759
Db 1494 LSLMENIKSQLEAKEIKQVDESTIQELDHEBITASKN---NYEGLKNDKDSII 1543

RESULT 9
B44450
ubiquitin-specific proteinase (EC 3.4.-.-) UBP3 - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YER151c; ubiquitin-specific processing proteinase
C/Species: Saccharomyces cerevisiae
C/Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 21-Jul-2000
C/Accession: B44450; S50654; S27467
R/Baker, R.T.; Tobias, J.W.; Varshavsky, A.
J. Biol. Chem. 267, 23364-23375, 1992
A/Title: Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of UBP2 and U
A/Reference number: A44450; MUID:93054674; PMID:1429680
A/Accession: B44450
A/Molecule type: DNA
A/Residues: 1-912 <BAK>
A/Cross-references: EMBL:M94917; NID:gl73129; PIDN:AAA35191.1; PID:gl73130
A/Note: sequence extracted from NCBI Backbone (NCBIP:118310)
R/Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A/Description: The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981, and lambda
A/Reference number: S50654
A/Accession: S50654
A/Molecule type: DNA
A/Residues: 1-912 <DIB>
A/Cross-references: EMBL:U18917; NID:G603377; PID:G603391; MIPS:YER151c
C/Genetics:
A/Gene: SGD:UBP3
A/Cross-references: SGD:S0000953; MIPS:YER151c
A/Map position: 5R
C/Keywords: hydrolase

```

```

Query Match 3.7%; Score 150; DB 2; Length 912;
Best Local Similarity 18.5%; Pred. No. 0.13;
Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;

QY 65 MTDAAARREGESLKKIQLPKSLTSLSSVSRGVNPPPHSHSSGLTPPTPTTPSSSSFRS 124
Db 3 MODANKEESYSYYPK-----TSSPPPTPTNMQIPIYQAPLQMYGTQ 45
QY 125 STPTETQI--VRELTRNLI--APIFLHR--LTYMSHRN-----SRTNLRKTF 169
Db 46 APYLVTQIIPAYSF---NMVNQNPYI-HQSGSPHILPPQNNINGSGTTNNNNKKW 100
QY 170 KVDDMLSKYERKMGQESHSLSAHLQLTFTGFHKNDK-----PSPS-----E 213
Db 101 HSGNITNN-NGSSNGQGANSSGSGMSYKSHYTHYNNHIPPMMASPSNGSNAGMKQT 159
QY 214 NEQN-----SVTLVLLVKVCHKRRKDVSCPIRQVPTGKKQVPLIDLNQT 259
Db 160 NSGNGSSATSPSYSSVNSQYDLYKFDVTKLKN---LKENSNLIQLPLF--INTT 213
QY 260 KPG-----NPPSLAVSNBEPFSN-----SHMVKSYSLFRVTRPGRREFNGM 302
Db 214 EAEFAAASVQRYELNMKALNLSSELSSESVKSSAHHHTKSHSI-----PKHNEE 264
QY 303 INGETN-ENIDVNEELPARKKRNREDGEKTFVAQMTVPDKRRRLQLLDGEYEVAMQEMEE 361
Db 265 VKTETHGEEDAHDKPHASKDAHELKKTEYK-----EDAKQDRNE 307
QY 362 CPISKKRAWETILDKRLPPETFSQGTFLQTLWTGETNDKSTAPIA---KPLATR 417
Db 308 KVIQEQATVLPVVKYE--PEESVEEN-----TSKTSPPSPSPAASWSAISAD 356
QY 418 NSESLHQENKPGSVKPTQTIIVKESLTTDLQTRKEKDTNENRQKLRIFYOFLYNNTRQ 477
Db 357 AIKRSOASNKTVSGSVVTKTPISGTAGVSTNMAAATIGKSSPL-----LSK 405
QY 478 QTEARDLHCPWCTLNCRKLYSLKHLKLCRGRFIPNYVYHPKGARIDVSNICEYDGSYA 537
Db 406 QQKQDKKVPSTKGIPLGSLA--LRMCPDPDFISVLRNKOVENKIPVHSII----- 458
QY 538 GNPQDIHQPGFAPSRNGPVKRTPIITHILVCRP-----KTKASM----- 577
Db 459 --PRGIINRANICF-----MSSVLQVLLYCKEFTDVINLSTRNTNGVGTSSCKLLDA 510
QY 578 -----SEFLESDG--EVEQORYSSGHNRLYFHSDTCLPLRQPEMEVDSE 621
Db 511 CLUTMYQFPKTEYKFFLENADDAKTTESDAKSSKSKSFQHCATADAVKPDF----- 565
QY 622 DEKDPWLREKTIITQIEFSDVNEGEKVMKLNVLNLMKH-----GFIADNMNHA 672
Db 566 -----YKLTSTIPKFKDLQWQHQAEBEFLTHLLDQLHEELISALDGLTONE---- 612
QY 673 CMLFVENYQKIIKKNLCRNFMLHLSVMH---DFNLISIMSIDKAVTKLRNQOKLEGE 729
Db 613 ----IQNMLQSIHQDL-KVFFIRNLSRYKAEF-----IKNASPRKLIELIYK---- 656
QY 730 SASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKOSKQK 775
Db 657 --GVINDDSTEB-----NGWHE-----VSGSKRGGKTK 683

```

```

RESULT 10
T18235
transcription activator GAL11 homolog - yeast (Candida albicans)
C/Species: Candida albicans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T18235
R/Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A/Reference number: Z18831
A/Accession: T18235
A/Status: preliminary; translated from GB/EMBL/DDBJ

```





A:Cross-references: EMBL:U35149; NID:gl049248; PID:gl049249; PIDN:AAA81320.1  
Query Match 3.6%; Score 147.5; DB 2; Length 1368;  
Best Local Similarity 18.7%; Pred. No. 0.33;  
Matches 174; Conservative 137; Mismatches 324; Indels 297; Gaps 46;

QY 33 TEDNH-----IDTPRVLEKQ---ELQOPTVALSYINRMTDAARQE--- 74  
DB 20 IDENHPLIKKHVHNLGTEHRVLYNYGKLPERYDYDSKNFVNYKTNNDSEEPDF 79  
QY 75 -----SLKKIKPKLSLTLSSSVSRGNVSTPRHSGSLTTPPTTPSSSFRSST 126  
DB 80 SNLFGSFEIPEKPEVKYNPFLGSK-----NNNSGS-----SFPSSFFNSDK 122  
QY 127 PTEPTQIYRFLRTRNLIAPIFLHRTLTVMSHRNRTNIRKRTFKVDDMLSKVEKMGQE 186  
DB 123 PTNS-----EILKDDNIEKNI---VDATVEDTESFNIRKREDNDAEKKVELLSLNGND 174  
QY 187 S-HSL-----AHLQLTFTGFFHKNDKPSNSEQNSVTLVLLV--K 227  
DB 175 SINLSINNIIDNTEHKVQVQNEQIEESKSFNDND---NSINDISS-----VLLIDNE 226  
QY 228 VCHKRDKVSCPIRQVPTGKQVQLIPDLNQTGPNFSLAVSNE--PEPS-----NSHM 281  
DB 227 TEHKVQVQNEQIEESKSFNDNDNDINDITENEKEH--KVQVQNEQIEESKSLSGNDS 284  
QY 282 VKSYSLFRVTPGRRFNGMNGETNENIDV--NEELPARKENRDEGKTFVAQVTF 339  
DB 285 INDISI-----NNVIDTEHKVQVQNEQIEESKSFNDNDNDINDISIMLLI 332  
QY 340 D-----KNRLQ-----LLDGEYEVAMQEMEECPISKKRATWETIL 375  
DB 333 DNTEHKVQVQNEQIEESKSLSGNDNDLNNVINEKEQIVQVQEKKSKRNM----- 387  
QY 376 DKRLPPETPSQGTLOF-----TLRWGTGTN-- 403  
DB 388 -----IFDEIRQGVIRKIKSSNNHKNMNIIVEKHEYLTOSEYEQKQKMLRNNRNVL 441  
QY 404 ----DKSTAPIAKPLATNSSLH--OEN-----KPGSVKPTOTIAVKSILT--TLOQT 449  
DB 442 FDLINKSNDL--KPYNEKH--EYIHVSEENEQIVKEIQALQPLQIKVNIKNMNFNITN 499  
QY 450 RKEKDTPTNENKQIRIFYQFLYNNNRQOTEARDLHCPWCTLNCRKLYS-----LLKHLK 505  
DB 500 NDTKDEYTDNNVESPNTQLYESNNEMYQNRG-----IYNCQIIFPDSVCIPKQE 550  
QY 506 LCHSRFIF-----NYVTHPKGARIDVSTNECYDGSYAGNPQDIHRQPGFAFRNGPV 557  
DB 551 YFHNEIYQALGIFPGSEVSPKQ---EYSYNDIYQLYGMTPNNEIYIPKGFENNEI 607  
QY 558 KRTP--ITHILVCRPK-----RTRKMSBEFLESEDEGEVEQ 591  
DB 608 YLSDMLSDCEVCTPKYBIEQNEEYIPLNGIFSDNEIYITEQKGLDKIEIYPLTGVSFN 667  
QY 592 RYSSGH-----NELYFHSOTCLPLR---PQEMEVDSEDEKDP 626  
DB 668 EYITEHIGSLNEEICPLNEVLLDNEIITEQIGSHDEEYIPLNEIPLNNEVITEQGS-- 725  
QY 627 EWLREKTTITQIEFSDVNEG-----EKEVMKLNWLVHMKGFADNMHACMLFV 677  
DB 726 --LDIEIYPLTGVSFD--NETYITEQKGLDKIEIYPL-----TGVSFSDNE-----IYIT 770  
QY 678 ENYGKQIKIKKLCRNFMHLVSMHDFNLIS--INSIDKAVTKLREMQQKLEKG-----ESAS 732  
DB 771 EHIGS--LNEEIC-----PLNEVLL--DNEIITEQIGSLDEAICPLNEILLDNEISTLEKDS 824  
QY 733 PANESITEQNGTANGFSEINSKKALETDSV 764  
DB 825 QDNEICTPKQEVSON--FEINTLNEILLSDNEI 854

TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.P  
N:Alternate names: protein P9659\_6  
C:Species: Saccharomyces cerevisiae  
C:Date: 09-Mar-1996 #sequence\_revision 06-Sep-1996 #text\_change 23-Mar-2001  
C:Accession: S69980  
R:Fulton, L.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of S. cerevisiae cosmid 9659.  
A:Reference number: S69022  
A:Accession: S69980  
A:Molecule type: DNA  
A:Residues: 1-1755 <FUL>  
A:Cross-references: EMBL:U40829  
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 41;  
C:Genetics:  
A:Map position: 16R  
A:Mobile element: retrotransposon Ty1.P  
C:Superfamily: TyB protein

Query Match 3.6%; Score 147.5; DB 2; Length 1755;  
Best Local Similarity 22.5%; Pred. No. 0.46;  
Matches 158; Conservative 98; Mismatches 234; Indels 213; Gaps 44;

QY 38 IDTPRVL---EKQELQOPTYVALSY---INRFMTD-----AAREQESLKKKI----- 80  
DB 881 VDTTNYVILQKGESRLDQPNYDALTFDEDLNRLTASYHSFIASNEIQESNDLINIESDHF 940  
QY 81 -----OPKLSLTLSSSVSRGNVSTPP-----RHSSGSLTPP--VTPPTTPS-- 119  
DB 941 QSDIELHPQPR--NVLSKAVSPD--STPSTHTSDSKRVSKTNIRASREVDPNISENI 997  
QY 120 --SSFRSPTPTQIYRFLTRN-----LIIPIFLHRTLTVMSHRNSRTIKRKT 168  
DB 998 LPSKKRSSTP---QISNIESTGSGMHKLVNPLAPMSQNT-----HESSHAS--KSKD 1047  
QY 169 FKVDDMLSKVEKMGQESHLSAHLQLTFTGFFHKNDKPSNPSENEQNSVTLVLLVKV 228  
DB 1048 FRHSDSYS-----ENETNHTNVPISSTG--GTNNKTVFQISDQETE-----KRI 1089  
QY 229 CHKRR--DVS-----CPIRQVPTGKKQ-----VPLIPDLNQTGKNFPLA 268  
DB 1090 IIRSPSIDASPPENSSHNIVPIKPTTVSEONTBESIIADLPL--PDLPPSPTEFPD-- 1146  
QY 269 VSSNEFEPSNSHMVKSYSLLFRVTPGRRFNGMNGETNENIDVNEELPARRKKNREDG 328  
DB 1147 -PFKELPPINSHTNS-----SLGGI--GDSNAYTTINS-----KKSLEBN 1185  
QY 329 EKTFF--VAQMTVFDKRRRLQLLDGEYEVAMQEMEECPISKRR---ATWETILDGRLPFP 383  
DB 1186 ETEIKVSRDTWNTKNR-----SLEPPRSKKRIHLIAVAV---KSIKPI 1228  
QY 384 ETFSGQPTLQF--TLRWGTGTNDSKSTAPIAKPLATNSSELSHQENKPGSVKPTQTIAYKE 441  
DB 1229 RT-----TLRYDEAITYNKDIKEK-----KYIEAYHKE-----VNQLLKMT 1266  
QY 442 SLTTDLQTRKEKDTNENRQKLIIFYOFLYNNNT-----ROQTEARDDLHCP-----WCTL 492  
DB 1267 WDTDKYDYDEKID-----PKRVNSMFIKRRKRGTHKARFVARGDIQHPDTYDTGMQS 1320  
QY 493 NCRKLYSLKHLKCHSRPFIYNYVHPKGARIDVSNIECYDGSYAGNPQDIHRQPGFAFS 552  
DB 1321 NTVHHVALMTSLSLDN---NYVI---TQLDIS--SAY--LYADIKEELVIRPPHUG 1369  
QY 553 RNPVGRPTPITHILVCRPKRTKASMEFLES---EDGEVQOQRTYSSGHNRLIFYHSDTCL 609  
DB 1370 MNDKLIRLKSHYGL---KQSGANWYETIKSYLIRKQCMEEVVRGWSG---VFKNSQVTI 1422  
QY 610 PLRPOEMEVDSEKDPWLREKTTITQIEEFSN-----VNEGEKE 649  
DB 1423 CLFVDDMLFSKDLN-----ANKKIITLKKQYDTKLIINLGSD 1461





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:57:34 ; Search time 138 Seconds  
(without alignments)  
1769.128 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 4078  
Sequence: 1 MTGIAAASFFSNTCRFGCG.....KALETDSVSGVSKSKQKL 776

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1297172 seqs, 314612898 residues

Total number of hits satisfying chosen parameters: 1297172

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4078	100.0	776	9	US-09-874-162A-8
2	3413	83.7	739	9	US-09-874-162A-5
3	2043	50.1	388	9	US-09-764-864-931
4	1497	36.7	289	9	US-09-764-864-911
5	1453	35.6	292	9	US-09-764-864-1369
6	919	22.5	175	9	US-09-764-864-1355
7	661	16.2	243	9	US-09-874-162A-2
8	661	16.2	243	15	US-10-144-198-16
9	661	16.2	243	15	US-10-104-047-2404
10	652	16.0	278	9	US-09-764-864-1334
11	476	11.7	143	9	US-09-764-864-884
12	359	8.8	69	14	US-10-029-386-28495
13	280	6.3	171	12	US-10-276-774-2060
14	214.5	5.3	813	14	US-10-231-778-2
15	211	5.2	692	14	US-10-177-478-6

ALIGNMENTS

RESULT 1

US-09-874-162A-8  
; Sequence 8, Application US/09874162A  
; Patent No. US20020155452A1  
; GENERAL INFORMATION:  
; APPLICANT: Koontz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZ1 GENES IN  
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874,162A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,093  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 776  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-162A-8

Query Match	100.0%;	Score	4078;	DB	9;	Length	776;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	776;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MTGIAAASFFSNTCRFGCGGLHPTLADLIEHEDNHIDTPRVLEKQELQQTYYVALSY	60				
Db	1	MTGIAAASFFSNTCRFGCGGLHPTLADLIEHEDNHIDTPRVLEKQELQQTYYVALSY	60				
Qy	61	INRMTDAARREQESLKKIKQPKLSLTSSVSNGVSTPPRHSSGSLTPTPTPTSS	120				
Db	61	INRMTDAARREQESLKKIKQPKLSLTSSVSNGVSTPPRHSSGSLTPTPTPTSS	120				
Qy	121	SFRSSTTEPTQIYRFLRNLNLAIPFLHRTLTVMHRSNRTNKRKTFKVDMLSKVEK	180				
Db	121	SFRSSTTEPTQIYRFLRNLNLAIPFLHRTLTVMHRSNRTNKRKTFKVDMLSKVEK	180				

QY 181 MKGEQSHLSAHLQITFTGPFHKNDKSPNSNEQNSVTLEVLVVKVCHKKRDXVSCPI 240  
Db 181 MKGEQSHLSAHLQITFTGPFHKNDKSPNSNEQNSVTLEVLVVKVCHKKRDXVSCPI 240  
QY 241 ROVPTGKQVPLIPDLNQTGKGNFPLAVSSNEFEPNSHVMVKSYSLLFRVTRGGRREFN 300  
Db 241 ROVPTGKQVPLIPDLNQTGKGNFPLAVSSNEFEPNSHVMVKSYSLLFRVTRGGRREFN 300  
QY 301 GMINGETHENIDVNEELPARKRNREDGKTFVAQMTVFDKRNRLQLLDGGEYVAMQEME 360  
Db 301 GMINGETHENIDVNEELPARKRNREDGKTFVAQMTVFDKRNRLQLLDGGEYVAMQEME 360  
QY 361 ECPISKKAATWETILDGKRLPPFETFSOGPTLQFTLRWTGETNDKSTAPIAKPLATRNSE 420  
Db 361 ECPISKKAATWETILDGKRLPPFETFSOGPTLQFTLRWTGETNDKSTAPIAKPLATRNSE 420  
QY 421 SLHQENKFGSVKPTQTIAVKSLSATDLOTRKEKOTPNENRQKLRIFYQFLYNNNTRQOTE 480  
Db 421 SLHQENKFGSVKPTQTIAVKSLSATDLOTRKEKOTPNENRQKLRIFYQFLYNNNTRQOTE 480  
QY 481 ABDLLHCPWCTLNCBKLKSLKHLKCHSRFIENYVHPKARIDVSNIECYDGSYAGNP 540  
Db 481 ABDLLHCPWCTLNCBKLKSLKHLKCHSRFIENYVHPKARIDVSNIECYDGSYAGNP 540  
QY 541 QDIHRQPGAFSRNGPVKRTPIHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHN 600  
Db 541 QDIHRQPGAFSRNGPVKRTPIHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHN 600  
QY 601 LYFHSDDCLPLRPQMEVDSDEKDPWLRKTTIQIEFSDVNEGEKEVMKLNHLHVMK 660  
Db 601 LYFHSDDCLPLRPQMEVDSDEKDPWLRKTTIQIEFSDVNEGEKEVMKLNHLHVMK 660  
QY 661 HGFIADNOMNHACMLFVENYQKIIKKNLCRNFMLHLSVMDHFNLSIMSDKAVTKLRE 720  
Db 661 HGFIADNOMNHACMLFVENYQKIIKKNLCRNFMLHLSVMDHFNLSIMSDKAVTKLRE 720  
QY 721 MQKLEKESASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776  
Db 721 MQKLEKESASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776

RESULT 2  
US-09-874-162A-5  
; Sequence 5, Application US/09874162A  
; Patent No. US20020155452A1  
; GENERAL INFORMATION:  
; APPLICANT: Koontz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JUAZ1 GENES IN  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874.162A  
; PRIOR FILING DATE: 2001-06-04  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 739  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-162A-5

Query Match 83.7%; Score 3413; DB 9; Length 739;  
Best Local Similarity 99.8%; Pred. No. 6.1e-282;  
Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 389 GPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKSLSLTTDLQ 448  
Db 389 GPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKSLSLTTDLQ 448  
QY 449 TRKEKOTPNENRQKLRIFYQFLYNNNTRQOTEARDLLHCPWCTLNCBKLKSLKHLKCH 508  
Db 449 TRKEKOTPNENRQKLRIFYQFLYNNNTRQOTEARDLLHCPWCTLNCBKLKSLKHLKCH 508  
QY 509 SRPIFYVHPKARIDVSNIECYDGSYAGNPDIHROPGAFSRNGPVKRTPIHILVLC 568  
Db 509 SRPIFYVHPKARIDVSNIECYDGSYAGNPDIHROPGAFSRNGPVKRTPIHILVLC 568  
QY 568 SRPIFYVHPKARIDVSNIECYDGSYAGNPDIHROPGAFSRNGPVKRTPIHILVLC 568  
Db 568 SRPIFYVHPKARIDVSNIECYDGSYAGNPDIHROPGAFSRNGPVKRTPIHILVLC 568

Query Match 50.1%; Score 2043; DB 9; Length 388;  
Best Local Similarity 99.5%; Pred. No. 2e-165;  
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## RESULT 3

US-09-764-864-931  
; Sequence 931, Application US/09764864  
; Patent No. US20020132753A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764.864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 931  
; LENGTH: 388  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-931

Query Match 50.1%; Score 2043; DB 9; Length 388;  
Best Local Similarity 99.5%; Pred. No. 2e-165;  
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 569 RPKTKASMESELESDGEVEQRTYSSGHNRLYFHSDTCLPLRPOEVEVDSDEKDEW 628
Db 181 RPKTKASMESELESDGEVEQRTYSSGHNRLYFHRDTCLPLRPOEVEVDSDEKDEW 240
Qy 629 LREKTIQTIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMHAHACMLFVENYQKIIKEN 688
Db 241 LREKTIQTIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMHAHACMLFVENYQKIIKEN 300
Qy 689 LCRNFMHLVSHMDFNLISIMSDKAVTKLRMOQKLEKESASPANEEITEEQNGTANG 748
Db 301 LCRNFMHLVSHMDFNLISIMSDKAVTKLRMOQKLEKESASPANEEITEEQNGTANG 360
Qy 749 FSEINSKEKALETDSVSGVSKQSKQKL 776
Db 361 FSEINSKEKALETDSVSGVSKQSKQKL 388

```

RESULT 4

```

US-09-764-864-911
; Sequence 911, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 911
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-911

```

Query Match 36.7%; Score 1497; DB 9; Length 289;

Best Local Similarity 99.6%; Pred. No. 5.2e-119;

Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy 493 NCRKLYSLKHLKLGHSRIENYVHPKGRADIVSINECYDGSYAGNQDTHRQGFAPS 552
Db 6 DCRKLYSLKHLKLGHSRIENYVHPKGRADIVSINECYDGSYAGNQDTHRQGFAPS 65
Qy 553 RNPVKRTPITHLVCRPKRTKASMESELESDGEVEQRTYSSGHNRLYFHSDTCLPLR 612
Db 66 RNPVKRTPITHLVCRPKRTKASMESELESDGEVEQRTYSSGHNRLYFHSDTCLPLR 125
Qy 613 POEMEVDSEKDPWLRKTTIQIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMHA 672
Db 126 POEMEVDSEKDPWLRKTTIQIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMHA 185
Qy 673 CMLFVENYQKIIKENLGNFNLHLSVHDFNLISIMSDKAVTKLRMOQKLEKESAS 732
Db 186 CMLFVENYQKIIKENLGNFNLHLSVHDFNLISIMSDKAVTKLRMOQKLEKESAS 245
Qy 733 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
Db 246 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 289

```

RESULT 5

```

US-09-764-864-1369
; Sequence 1369, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper

```

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1369

LENGTH: 292

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (9)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (16)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (18)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (31)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (46)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (121)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (283)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-864-1369

Query Match 35.6%; Score 1453; DB 9; Length 292;

Best Local Similarity 96.8%; Pred. No. 3e-115;

Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 233 RKDVSCPIROVPTGKQVPLIPDLNOKPGNFFPSLAVSSNEFSPNSHVMKYSLLPRVT 292

Db 1 KDVSCPIXQVPTGKQXPLNPLNQLNQTGKXFFSLAVSSNEFSPNSHVMKYSLLPRVT 60

Qy 293 RPSRRFNGMNGETNENIDVNEELPARRKRNEDGKTFVAQMTVPDKNRRLQLLDGEY 352

Db 61 RPSRRFNGMNGETNENIDVNEELPARRKRNEDGKTFVAQMTVPDKNRRLQLLDGEY 120

Qy 353 EVANQEMEECPISKRRATWETILDGKELPPFETFSQGPITQTLRTWGTENDKSTAPIAK 412

Db 121 XVANQEMEECPISKRRATWETILDGKELPPFETFSQGPITQTLRTWGTENDKSTAPIAK 180

Qy 413 PLATRNSESLHQENKPGSVKPTQTIAVKESLITDLOTRKEKDTPNENRQKLRIFYQFLYN 472

Db 181 PLATRNSESLHQENKPGSVKPTQTIAVKESLITDLOTRKEKDTPNENRQKLRIFYQFLYN 240

Qy 473 NNTQQTEARDDHLCPWCTLNCRKLYSLKHLKCHSRFTFNXY 516

Db 241 NNTQQTEARDDHLCPWCTLNCRKLYSLKHLKCHSRFTFNXY 284

RESULT 6

```

US-09-764-864-1355
; Sequence 1355, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1355
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

; NAME/KEY: SITE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (136)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (166)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-864-1355

Query Match 22.5%; Score 919; DB 9; Length 175;  
Best Local Similarity 94.3%; Pred. No. 5.2e-70;  
Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 466 FYQFLNNNTROQTARDLHCPCWCTLNCRLKLSLLKHLKLSRFFINYYHHPKGARID 525  
DB 1 FYQFLNNNTROQTARDLHCPCWCTLNCRLKLSLLKHLKLSRFFINYYHHPKGARID 60  
QY 526 VSINECYDGSVAGNPDHROPQFAPSRGVPKRTPIITHILVCRPKTKASMSFEFLESED 585  
DB 61 VSINECYDGSVAGNPDHROPQFAPSRGVPKRTPIITHILVCRPKTKASMSFEFLESED 120  
QY 586 GEVEQORTYSSGHNRLYFHSOTCLPLRPOEVEVSEDEKDEWLRKXTITTOIEEF 640  
DB 121 GEVEQORTYSSGHNRLYFHSOTCLPLRPOEVEVSEDEKDEWLRKXTITTOIEEF 175

RESULT 7  
US-09-874-162A-2  
; Sequence 2, Application US/09874162A  
; Patent No. US20020155452A1  
; GENERAL INFORMATION:  
; APPLICANT: Koontz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZ1 GENES IN  
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874,162A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,093  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-874-162A-2

Query Match 16.2%; Score 661; DB 9; Length 243;  
Best Local Similarity 100.0%; Pred. No. 9.1e-48;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTGIAAASFNTCRFGGCGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTVVALSY 60  
DB 1 MTGIAAASFNTCRFGGCGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTVVALSY 60  
QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120  
DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120  
QY 121 SFRSSTPT 128  
DB 121 SFRSSTPT 128

RESULT 8  
US-10-144-198-16  
; Sequence 16, Application US/10144198  
; Publication No. US20030219748A1  
; GENERAL INFORMATION:  
; APPLICANT: OriGene Technologies Inc

; TITLE OF INVENTION: Regulated Prostate Cance Genes  
; FILE REFERENCE: 9U105 R1  
; CURRENT APPLICATION NUMBER: US/10/144,198  
; CURRENT FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 16  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-144-198-16

Query Match 16.2%; Score 661; DB 15; Length 243;  
Best Local Similarity 100.0%; Pred. No. 9.1e-48;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTGIAAASFNTCRFGGCGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTVVALSY 60  
DB 1 MTGIAAASFNTCRFGGCGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTVVALSY 60  
QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120  
DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120  
QY 121 SFRSSTPT 128  
DB 121 SFRSSTPT 128

RESULT 9  
US-10-104-047-2404  
; Sequence 2404, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1el full length cdna  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2404  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2404

Query Match 16.2%; Score 661; DB 15; Length 243;  
Best Local Similarity 100.0%; Pred. No. 9.1e-48;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTGIAAASFNTCRFGGCGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTVVALSY 60  
DB 1 MTGIAAASFNTCRFGGCGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTVVALSY 60  
QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120  
DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120  
QY 121 SFRSSTPT 128  
DB 121 SFRSSTPT 128

RESULT 10  
US-09-764-864-1334  
; Sequence 1334, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies



```

; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1334
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1334

Query Match      16.0%; Score 652; DB 9; Length 278;
Best Local Similarity 99.2%; Pred. No. 6.6e-47;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDRVLEKQELQOPTVALSY 60
DB 36 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDRVLEKQELQOPTVALSY 95
QY 61 INRFMTDAARRQESLKKIKQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPITPSS 120
DB 96 INRFMTDAARRQESLKKIKQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPITPSS 155
QY 121 SPFSSTPT 128
DB 156 SPFSSTPT 163

RESULT 11
US-09-764-864-884
; Sequence 884, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 884
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-884

Query Match      11.7%; Score 476; DB 9; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.5e-32;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDRVLEKQELQOPTVALSY 60
DB 51 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTDRVLEKQELQOPTVALSY 110
QY 61 INRFMTDAARRQESLKKIKQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPITPSS 93
DB 111 INRFMTDAARRQESLKKIKQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPITPSS 143

RESULT 12
US-10-029-386-28495
; Sequence 28495, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

```

```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28495
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHRI7.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P23477, EVALUATE 2.70e+00
US-10-029-386-28495

Query Match      8.8%; Score 359; DB 14; Length 69;
Best Local Similarity 98.6%; Pred. No. 7.9e-23;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 569 RPXETKASMSEFLESEDEGEVQQTYSXGHNRLYFHSDTCLPLRPQMEYDSEKDPW 628
DB 1 RPXETKASMSEFLESEDEGEVQQTYSXGHNRLYFHSDTCLPLRPQMEYDSEKDPW 60
QY 629 LREXITITQI 637
DB 61 LREXITITVI 69

RESULT 13
US-10-276-774-2060
; Sequence 2060, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2060
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2060

Query Match      6.9%; Score 280; DB 12; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.8e-15;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 QESLKKIKQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPITPSSPSSTPT 128
DB 1 QESLKKIKQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPITPSSPSSTPT 56

RESULT 14
US-10-231-778-2
; Sequence 2, Application US/10231778
; Publication No. US20030126647A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau, Pierre
; APPLICANT: Chaudhury, Abdul M.

```

```
Query Match      5.3%; Score 214.5; DB 14; Length 813;
Best Local Similarity 18.5%; Pred. No. 7.6e-09;
Matches 147; Conservative 113; Mismatches 298; Indels 237; Gaps 28;
```

---

```
QY      129 EPTQVRLFRNLNLAIFLHRTLYK--SHRSRNTNIKK-TFKVDDMLSKVERMKGEQ 185
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       70 KPVAVYKLETSKNPNPFLRSKIYLIAQKKKKNSGGKIRFYRVDSNKWTUKAEVV 129
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      186 ESHS-----LSAHL-----194
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       130 ENFSCPFCLIPCGGHEGLHLKSHDAFFEFYRAEKDHGPEVDVSVKSDTIKFGVLKD 189
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      195 -----QLTFTGFHHNDKPSPNSEONSVTLEVLIVKV-----CHKRRXDV 236
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       190 DVGNPQLSPLFCSKQRNQRRORDSNNVKLNVLLMELDDLDPGTENDSTHVNDNV 249
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      237 SCPIR-----QVPTCKKQVPLIPLDNQTKPGNF--PSLAVSSNE-----273
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       250 SSPPPRAHSSEKISDILATTLQTAIAESSEPVPKHVD---GNVSSPPRAHSAEKNESHV 306
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      274 -----PPSPNSHMVKSYSLLF---RVTPGRBRENGMINGETWENIDVNEE---LPAR 320
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       307 NDDDVDVSPPPRAHSLKKNESHVNEDNISPPKAH-----SSKQNESTHMDEDDVSFPFR 361
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      321 RKRNRDEGETFVAQMTPDK-----NRRLLQLLDGEYVAMQEMEECPISKRAWE 372
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       362 TRSSKETSDILTTQTPAIVEPSEPVPKVRGSRKQLYAKY-----KARETOP 408
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      373 TILDGKRLPPF---EIFSQGP-----TLQTLRWTCF---TNDKSTAPI 410
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       409 AIAESSEPVLHVNDENVSPPPRAHSLKASDILTTQTAIAESSEPVPKHVNDENVST 469
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      411 AKPIATRNSLSHQ--ENKPGSVKPTQTIAVKESLTDTLOTRKEKDTP-----NNNRQK 462
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       469 PRAHSSKKNKSTRKNVDNVPSPPKTRSSKKTSDILTITQTAIESSEPVPKHVNDENVSS 528
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      463 LRIYFQFLYNNTRQTEARDLHLCPCWCTLNCRKLYSLKLKLCBSAFIFYNYTHPKGA 522
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       529 TPRAHSSKKNKSTRND---DNTPSPPKTRSSKKNILTRIQQ-----PAIA 572
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      523 RIDVSINECYDGSVAGNP-----QDIHRQPGFA-----FSRNGPVKR 559
       ||::||::||::||::||::||::||::||::||::||::||::||::||:
```

QY 341 KNRRLQLLDGEYEVAMQWEECPISKKRATWETILDGKRLPPFFTFSGPTLQFTLRWTG 400  
Db : : : : :  
QY 392 SS-----EPKRVHVDNDVSS-----TPRAHS 413  
Db : : : : :  
QY 401 ETNDKSTAFIAKPLATRNSESJHQENKPGSVKPTQTIAVKSLSLTTDLQTRKEKDTNENR 460  
Db : : : : :  
QY 414 SKNKST-----RKDDNIPSPPKTRSSKKTNLTTR---TOPAIAESEPKVPHVND 462  
Db : : : : :  
QY 461 QKLRIFYOFLYNNTRQCTEARD--LHCPWCTLNCRKLYSLKHLKLCCHSRFIFNYVYHP 519  
Db : : : : :  
QY 463 DKVSSTPR-AHSSKKNKSTHKDDNASLPPKTRSSKKTSDIL----- 503  
Db : : : : :  
QY 520 KGARIDVSIKCYDGSYAGNPQDIHRQPGAFSRNGPVKRTPTITHILVCRPKRTKASME 579  
Db : : : : :  
QY 504 -----ATTQPAKAP-----SEPKVTRVSRK 525  
Db : : : : :  
QY 580 FLESDGEVEQOQTYSSGHNRLYPHSDTCLPLRPOEM--EVDSEDEKDPWLREKTIITQI 637  
Db : : : : :  
QY 526 ELHAERCEAKRLERLK---GRQFYHSQTMQPMTFEQVMNEDSENETDDYALDISERLRL 582  
Db : : : : :  
QY 638 BEFSDVNEGEKEVMKLNILHVMKHGFIADNQMHACMLFVE-----NYGQKLI 685  
Db : : : : :  
QY 583 ERLVGVSKEEKYMYLWNIIFVKQKRVADGHVPWACEBFKHLKEEMKNSSSFDWWWMF 642  
Db : : : : :  
QY 686 KKNLCRNFMLHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITE 740  
Db : : : : :  
QY 643 RIKLWNGLICAKTFHKCTTILLSNDEA-----GQFTSGSAANANNOQSME 689  
Db : : : : :

Search completed: August 25, 2004, 18:09:35  
Job time : 148 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:53:09 ; Search time 34 Seconds  
(without alignments)  
1178.287 Million cell updates/sec

Title: US-09-874-162A-8  
Perfect score: 4078  
Sequence: 1 MTGIAAASFFNTCRFGCGG.....KALETDSVSGVSKSKQKL 776

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/protdata/2/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/protdata/2/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/protdata/2/iaa/5A-COMB.pep:\*  
4: /cgn2\_6/protdata/2/iaa/5B-COMB.pep:\*  
5: /cgn2\_6/protdata/2/iaa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/protdata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238.5	5.8	604	US-09-339-947A-8	Sequence 8, Appli
2	224	5.5	611	US-09-339-947A-1	Sequence 1, Appli
3	159	3.9	1494	US-08-755-587-186	Sequence 186, App
4	156.5	3.8	3696	US-09-134-001C-5080	Sequence 5080, Ap
5	150	3.7	912	US-07-789-915A-8	Sequence 8, Appli
6	150	3.7	912	US-08-005-002C-8	Sequence 8, Appli
7	150	3.7	912	US-08-487-203A-8	Sequence 8, Appli
8	146.5	3.6	1235	US-08-118-101A-2	Sequence 2, Appli
9	142.5	3.5	1972	US-09-418-710-21	Sequence 21, Appli
10	137.5	3.4	2954	US-09-150-867-1	Sequence 1, Appli
11	135	3.3	3248	US-08-353-700-1	Sequence 1, Appli
12	135	3.3	3248	PCT-US95-16216-1	Sequence 1, Appli
13	132	3.2	2482	US-08-328-254-6	Sequence 6, Appli
14	128.5	3.2	907	US-08-938-830-26	Sequence 26, Appli
15	128.5	3.2	907	US-09-020-222-26	Sequence 26, App
16	128.5	3.2	1167	US-09-803-671B-2	Sequence 2, Appli
17	128	3.1	798	US-09-203-453-5	Sequence 5, Appli
18	128	3.1	798	US-09-900-236-5	Sequence 5, Appli
19	128	3.1	1164	US-09-457-708-2	Sequence 2, Appli
20	128	3.1	1164	US-09-950-046A-2	Sequence 2, Appli
21	128	3.1	1164	US-09-976-594-989	Sequence 989, App
22	127.5	3.1	609	US-09-107-532A-4451	Sequence 4451, Ap
23	126.5	3.1	3418	US-08-639-501-2	Sequence 2, Appli
24	126.5	3.1	3418	US-08-603-753D-4	Sequence 4, Appli
25	126.5	3.1	3418	US-09-044-946-2	Sequence 2, Appli
26	126.5	3.1	3418	US-08-755-587-44	Sequence 44, Appli
27	126.5	3.1	3418	US-09-044-908-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-339-947A-8  
; Sequence 8, Application US/09339947A  
; Patent No. 6630616  
; GENERAL INFORMATION:  
; APPLICANT: YOSHIDA, No. 6630616umasa  
; APPLICANT: YANAI, Yukihiko  
; APPLICANT: KATO, Yoshihiro  
; APPLICANT: HIRATSUKA, Junzo  
; APPLICANT: TAKAHASHI, Shigeru  
; APPLICANT: MIWA, Tatsuichi  
; TITLE OF INVENTION: GENE FOR FLORAL REGULATION AND METHODS FOR CONTROLLING  
; FILE REFERENCE: 032735-001  
; CURRENT APPLICATION NUMBER: US/09/339,947A  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: JP 1998-180065  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: JP 1999-179043  
; PRIOR FILING DATE: 1999-06-24  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 604  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: ZN FING  
; LOCATION: (310)..(335)  
US-09-339-947A-8

Query Match 5.8%; Score 238.5; DB 4; Length 604;  
Best Local Similarity 18.9%; Pred. No. 2.5e-13;  
Matches 130; Conservative 95; Mismatches 242; Indels 221; Gaps 22;

QY	129	BPTQIVRFLRNRNLIADIFLHRTLTVMSH--RNSRTNI-----KKTPTKVDMDLSKVE	179
DB	27	KPVLYNIIQERSIKNPAFLQRCCLLYKHARKKRSLLITISLSGGTNKELRAQNIFFLYV	86
QY	190	KMKGEQSHSLSAH-----LQLTFTGFFHNDKPS-----PNSGN--EQNSVTLEV	223
DB	87	LLARPNTNVSLEGHSPYRFRSACLLTSFHEFGNKDYTEATFVDPVKNLATSACSLNI	146
QY	224	LLVKVCHKKK--DVSCPIRQVPTGKKQ-----VPLIDLNQTKPGNFPPLSVS	270
DB	147	ILIS-CGRABOTFDNNCSGNHVEGSTLQLEKGCFKGKIFI--DLLASSGNCVLSL-	202
QY	271	SNEPEPSNHWKSYSLFRVTRPGRREFNGMNGTNEIDVNEELPA--RRKRNREDG	328
DB	203	-----GHTVEMSGSTVEMTPTSFLEPKFLEDD	228



```
/ APPLICATION NUMBER: GB 9523959.6
/ FILING DATE: 23-NOV-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: GB 9525555.0
/ FILING DATE: 14-DEC-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: GB 9617961.9
/ FILING DATE: 28-AUG-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kenneth D Sibley
/ REGISTRATION NUMBER: 31,665
/ REFERENCE/DOCKET NUMBER: 5405-135
/ INFORMATION FOR SEQ ID NO: 186:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1494 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-755-587-186

Query Match      3.9%; Score 159; DB 3; Length 1494;
Best Local Similarity 18.3%; Pred. No. 3e-05;
Matches 177; Conservative 151; Mismatches 355; Indels 284; Gaps 46;

QY 23 FPTLADLIEHEDNHDTPRVLEKQELQOPTVVA-LSVINRFMTDAARREOESLKKIKI 81
DB 370 FTKASKEIKLSENNIRKSMFKDIEHYPTNLACLEIVNTSSLESQKKPSK--HALD 427

QY 82 PKLSLTSSVSRGNVSTPPRHSSGSLTPPTPPITPSSSRSSSTPTPTQI----- 133
DB 428 PQINI-IGSFVQ--NSTVYDSGSHGTAPTLSLQDPDSNNLTPSQKABITELSTILE 484

QY 134 -----YRPTLRNLAPIELHRTLTYMSHNSRT--NIKRTFKVDMM-----LSKVE 179
DB 485 ESGSQEFTQFRK---PSHIIQKNPEMPENOLTILNSTSEKWKDDDLHTNAPSIQVD 541

QY 180 KMK-----GEQSHSLSA-----HLQLTFTGFPH-KNDKPSPNSE----- 213
DB 542 SKKSEGIIGKQKPAFLSRTSCNRASGYSTDKNEVEFRGFYSARGTKLVGSEALQXAK 601

QY 214 -----NEONSV-----TLEVLVVKVCHKKKVDKSCPIRQVPTGKKQP 251
DB 602 KLFSDLNENIETSEVDRFSSSKYNDVSMVQIEDCNDKNE-----PNKKRL- 652

QY 252 LIPDLNQTGPNFPPSLAVSSNEFEPSPNSHWKYSILLR-----VTRPGREFNGM 302
DB 653 -----ILQNIEMTDFVBEYTESYRRNTENEGNQCTDAGRNTCSSE 695

QY 303 ING---ETNENIDVNEE---LPARRKRNREDEKTFVAQMTVFDKNRLQLLDGEYE--- 353
DB 696 SDGSDSSKNDTVYIHEENGPLCIDQHNID-----LKLFSQFMKEGNTQIKEGLSDUTC 749

QY 354 VAMQMEECPISKKRATWETILDGKLPPEFTE--SQGPTLQFTLRWTGETNDKSTAPIA 411
DB 750 LVNKABEETSHVTMSNKKQJLTANTGQNIKDFDTFYLSFQTASRKNIRVSESLNKARSLLN 809

QY 412 KPLATRN---GESLHOENKPG-SVKPTQTIIVAKESLTTDLQTRKEKDT----- 455
DB 810 QKWTEBELNFSDSLNSSELLPGDIDIKTD---ISNHEVIENTERKDKITKESLIGTENIL 866

QY 456 -----PNENRQKRLFYOFLYNNNTROQTEARDDLHCPCWTLNCRKLYSLKHLKLS 509
DB 867 LILQORPESKIKIKESAVLGFHTASGKKIETKE-----SLDKVKNLFEKEQDMS 918

QY 510 REIFNVYVHPKGARIDVSNIECYDG-SYAGNPQDIHRQPGFAFRNGPVKRPITPILVLC 568
DB 919 E-ITNFSH--RGAKMSKDRBCKDGLACGTEITTTTEYETSHSSLEKKVSVSEIALAL 975

QY 569 RP-----KRT-----KASMEFEILEDGE-----VEQORTYSSGHNH-LYFH 604
DB 976 RPRLSLDNLKYQENLKI SDHASQKVDVHENTEKETAKKPTVMYTNQSTYSAIENSPLATF 1035

QY 605 SDTCLPLRQMEVDSEDE--KDPFWLREKTIQIIEFSD-----VNEGEKEVWKLWNLHV 658
```

```
DB 1036 QDT-----BEKFSVSEASLFEAKWMLRE-----GEWDDQSERINAACKVCKEYDDYV 1083
QY 659 -----MKHGFTADNQNHACHMLFVENYQKIKKNLCRN-FWLHLVSMHD---FNLIS 707
DB 1084 ENPSCGNSNSAITENDKNH-----LSEKQGSTVLSNSTMSNSYHPGFCSSSEVYNKSE 1139
QY 708 IMS-----IDKAVTKLREMQ-----OKLEKGESASP 733
DB 1140 YLSRSKIDNSGIEPVIKNIERKNIGPSEIMSPGREADTDPQSVNDICVEKLATNSCKN 1199
QY 734 AN-----BEITEQGTANGFSEINSKE-----KALS--TDSVSGVS 768
DB 1200 KNTAKVAISDSNNFNITQKLNDSNNSVPAYSTVNSKRVFAHQTKVTSFGTDNCMSVT 1259
QY 769 QSKXKQK 775
DB 1260 KQNTKSK 1266

RESULT 4
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynda Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match      3.8%; Score 156.5; DB 4; Length 3696;
Best Local Similarity 17.4%; Pred. No. 0.00022;
Matches 140; Conservative 147; Mismatches 327; Indels 189; Gaps 32;

QY 30 IEHIEDNHDTPRVLEKQELQOPTVYVALSVINRFMTDAARREOESLKKIKPQLSLTSL 89
DB 1926 LKINNIPQSTQTKNAKQEIINDKAQQLIOINN--TPDATEEKEQEAATNRVNAQLAQAIQ 1984
QY 90 SSVSRGNVSTPPRHSSGSLTPPTPTPSSSFSSSTPTPTQIYRFLRTRNLIAPIFLH 149
DB 1985 -----NINN--AHSTQVNESKTNSTATIKSVQPNVIKKTAINSLTQGAN----- 2028
QY 150 RLTLYMGRHSRTNIKK-----TFKVDMLSKVKMKKGQESHSLSAH---LQLTFT 199
DB 2029 NQKTLIGNDGNATDDEKAAKQLVTQKLEQIKIHESQDNQVDNKAQAITAIKLI-- 2086
QY 200 GFPHKNDKPSNSEQNSVTEVLVVKVCHKKKVDKSCPIRQVPTGKQVPLIPDLNQT 259
DB 2087 -----NANAKRQDAIN---ILTNLAESKASDIRA--NQDATTEKNATQSIDDT 2132
QY 260 KPGNFPSLAVSSNEFEPSPNSHWKYSILLFRVTPRGRREFNGMTG----- 305
DB 2133 -----LAQARNNINGANTNALVDENL-----EDGKQKLQRIVLSTQTKQAKADIAQA 2180
QY 306 -----ETNENIDVNEELPARKNRE-----DGEKTFVAQMTVFD-KNRLQLLDGE 351
DB 2181 IGOORSTIDQONATTEKOBALERLNCQETNGVNDRIQAALANQNVTDKKNILLETIRNV 2240
QY 352 YEVAQMEECPISKKRATWETILDGKRLPPFPFSSQPTL---QFTLRWTGETNDKSTA 408
DB 2241 EPIVIVPKANEILIRKAAEQTLINQ-----NQDATLEEKQIALGLEEVKNEALN 2292
```

QY 409 PIKPIATNSSESLQEN-----KPSGVKPTOTIA--VKESLTDTLOTRKEDTNNRQK 462  
 DB 2293 QVSO--AHNSNDVKIAENNGIAKISEVHPETIIKENAKQIEBDAQS--QIDTINAN---2345  
 QY 463 LRIYQFLYNNTRQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIFYVYHPKGA 522  
 DB 2346 -----NKSTNEEKSNAID-----RVNVAKIDAI-----NNITNATT 2377  
 QY 523 RIDVINECYDSYAGNPODIHQPCFAPSRNGPVKRTPIITHILVCRPKTKASSEFLE 582  
 DB 2378 QL---VN---DAKNGNTSISQILPSTA-----VK-----TNALAALASEAKNNAIIDQ 2421  
 QY 583 SEDGEVEQRTYSSGHNRLYFHSIDTCLPLRPOBMEVDEDEKDFEWLREKTIQI-----637  
 DB 2422 TPNATAEKEEBAANNKVDRL-----QEEADANILKAHTTDEVNNIKN 2462  
 QY 638 BEFSDVNEGEKEVMKMLNHLVHMKGFIADQNMHACMLFVENYQKIIKXLCRNFMHL 697  
 DB 2463 QAVQINAVQVEVIKKQNVQNLQFI-DNOKK-----IIENTPDATLEEKAEANRLIQN 2516  
 QY 698 V-----SMHDFNLISIMSDIKAVTKLREMOQKLEKESGASPANEEITSEQNGTANGF 749  
 DB 2517 VLTSTSDETIANDHNEVDQALDKARPKIEEIVPQVSKRDVLAIOBAFNSQOEIQEN 2576  
 QY 750 SEINSKKALETDSVSGVSKQSK 772  
 DB 2577 QEATNEEKTEALNKINQLLNQAK 2599

RESULT 5

US-07-789-915A-8  
 ; Sequence 8, Application US/07789915A  
 ; Patent No. 5212058  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Rohan T.  
 ; APPLICANT: Tobias, John W.  
 ; APPLICANT: Varshavsky, Alexander  
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/789,915A  
 ; FILING DATE: 19911108  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brook, David E.  
 ; REGISTRATION NUMBER: 22,592  
 ; REFERENCE/DOCKET NUMBER: MIT-509AA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 912 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-789-915A-8  
 ; Query Match 3.7%; Score 150; DB 1; Length 912;  
 ; Best Local Similarity 18.5%; Pred. No. 9.4e-05;  
 ; Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;

QY 65 MTDAAAREOESLKKIQPKLSLTSSSVGRGNVSTPPRHSSGSLTPPVTPPTTPSSSFRS 124  
 DB 3 MQDANKEESVMPK-----TSSPPPTPTNQIPIYQAPLQVMGYTQ 45  
 QY 125 STFTBTOI--YRFLTRNLI--APIFLHRT-----LYTMSHRN-----STNKKRKT 169  
 DB 46 APYLYTQIPAYSF-----NMVQNQPIY-HQSGSPHLPPONNINGSGSTNNNNKKW 100  
 QY 170 KVDMLSKVEKMKGESEHSLSAHLQLTFTGFHKNDK-----PSPNS-----E 213  
 DB 101 HSGIINN-NGSGNGGANGSGSGSYNKSHTYHNYNNHIFPMASPNSSGSMAGMKOT 159  
 QY 214 NEON-----SVTLEVLIVKCHKRDVSCPIRQVPTGKQVPLIPDLNQ 259  
 DB 160 NSSNGSGSATSPSYSSYSSSQDYLYKFDVTKLN-----LKENGSNLIQLPLF--INTT 213  
 QY 260 KPG-----NFPSLAVSNEPEPSN-----SHMVKSYSLLFRVTRPGRRENGM 302  
 DB 214 EAFPAASVORVELNKNALNLSSELSSEKSAHHHTKSHS1-----PAHNEE 264  
 QY 303 INGETN-ENIDVNEELPARKENREDGKTFVAQMTVFDKRRRLQLLDGEYEVAMQEMBE 361  
 DB 265 VKTETHGESEDAHDKKPHASKDAHELKKKTEVKK-----EDAKQDRNE 307  
 QY 362 CPISKKRATWETILDGKRLPPPETPSOGTLOFTLRWTCETNDKSTAPIA-----KPLATR 417  
 DB 308 KVIQEPQATVLPVVDKKE--PEESVEEN-----TSKTSPPSPFPAKSAWSAIASD 356  
 QY 418 NSESLHQENKPGSVKPTQTIKVESLTDLQTRKEKDTPENENRQKLRIFYQFLYNNTRQ 477  
 DB 357 AIKSRQASNKTVSGSMVTKTPISGITAGVSSINMAAATIGKSSSL-----LSK 405  
 QY 478 QTEARDDLHCPWCTLNCRKLYSLKHLKCHSRFIFYVYHPKGARIDVINECYDGSYA 537  
 DB 406 QPKQDKKYVPFSTKIEPLGSA--LRMCFDPDFISYVLRNKOVENKIPVHSII-----458  
 QY 538 GNPQDIHROPGFAGSRNGPVKPTPIHILVCRP-----KRTKASM-----577  
 DB 459 --PRGIINRANTCF-----MSSVLQVLLYKCPFDIVNLSTRNTRNGVTSCKLIDA 510  
 QY 578 -----SEFLESDG--EVSQORTYSSGHNRLYFHSIDTCLPLRQPMVEVDSE 621  
 DB 511 CLTMWKQFDKETYEKKFLENADDAKTTESDAKSKSKSFOHCATADAVKPDF-----565  
 QY 622 DEKDPEWLEKTIQIEBESDVNEGEKEVMKMLNHLVHMKG-----GFIADQNMHA 672  
 DB 566 -----YKTLSTIPKFDLQWGHQDABEFLTHLLDQLHEELISAIDGLTDNE-----612  
 QY 673 CMLFVENYQKIIKKNLCNFMHLVSMH---DFNLISIMSDIKAVTKLREMOQKLEGE 729  
 DB 613 ----IQNMLQSIINDEQL-KVFFIRNLSRYGKAEP-----IKNASPRKELIEXY----656  
 QY 730 SASPANEETIETONGTANGFSPINSKEKALETDSVSGVSKQSKKQK 775  
 DB 657 --GVINDDSTEE-----NGWHE-----VSGSSKRGKTK 683

RESULT 6

US-08-005-002C-8  
 ; Sequence 8, Application US/08005002C  
 ; Patent No. 5494818  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Rohan T.  
 ; APPLICANT: Tobias, John W.  
 ; APPLICANT: Varshavsky, Alexander  
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kevin M. Farrell, P.C.  
 ; STREET: P.O. Box 999  
 ; CITY: York Harbor  
 ; STATE: Maine



```

; COUNTRY: U.S.A.
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/005,002C
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,915
; FILING DATE: 08-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: MIT-5091AAZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-005-002C-8

Query Match 3.7%; Score 150; DB 1; Length 912;
Best Local Similarity 18.5%; Pred. No. 9.4e-05;
Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;

QY 65 MTDAAAREQESLKKIKQPKLSLTSSVSRRGNVSTPPRHSSGSLTPPVTPPTPSSSFRS 124
Db 3 MQDANKESYSVYK-----TSSPPPTPTNNQIPIYQAPLQWGYTQ 45

QY 125 STPTPEPTQI--YRFLRTRNLI---APFLHRT-----LTYMSHRN-----SRTNKRKTF 169
Db 46 APYLYPTQIYAYSPF---NMVNQNQPIY-HQSGSPHLPFPQNNINGSGSTNNNNKKKW 100

QY 170 KYDDMLSKYKMKGEQSHLSAHLQLTFTGFHNDK-----PSPNS-----E 213
Db 101 HSGNTNN-NGSSNGGANSNGSGSYNKSHTYHNYNNHIPPMAVSPNSGNAGKKQT 159

QY 214 NEQN-----SVTLEVLVVKCHKRQDVCPVIRQVPTGKKQVPLPDLNQT 259
Db 160 NSSNGSGSATSPSYSSVNSQYDLYKPDVTKLKN---LKENGSNLIQLPLF--INTT 213

QY 260 KPG-----NPSLAVSNSEFEPN-----SHVVKSYLLFVTRPGRREFNGM 302
Db 214 EAEFAAASVORVELANKALNLSSELSNSVEKSSAAHHTKSHSI-----PKHNEE 264

QY 303 INGETN-ENIDVNEELPARKENREDGEKTFVAQMTVFDKNNRRLQLLDGEYEVAMQEMEE 361
Db 265 VXTETHEEEDAHDKPHASKDAHELKKKTEYK-----EDAKQDRNE 307

QY 362 CPISKRAFWETILQKRLPPETSPQGTLOFTLRWGETNDKSTAPIA-----KPLATR 417
Db 308 KVIOEPQATVLPVDKKE--PEESVEEN-----TSKTSPPSPSPAASWGAIASD 356

QY 418 NSESLHQENKPGSVKPTQIATKVESLTTDLQTRKEKDTNENRQKRLIFYQFLYNNTRQ 477
Db 357 AIKRSQASNKTVSGMWVTPISGTTAGVSSSTNNAAATIGKSSPL-----LSK 405

QY 478 QTEARDDLCPMCTLCNCRKLYSLKLKELKLCCHSRFIFNVYVYHPKGARIDVSNICEYDGSYA 537
Db 406 QPOKDKKVVPPSTKGIEPLGSGIA--LRMCFDPDFISVYLRNKNVKNIPVHSII----- 458

QY 538 GNPQDIHRQPGFAFSENGPVKPTPIITHILVCP-----KPTKSM----- 577
Db 459 ---PRGIINRANTCF-----MSSVLQVLLYCKEFDIVINVLSTRNTNSRGTSSCKLLDA 510

```

```

QY 578 -----SFELESDG--EVEQORTSYSSGHNRLYFHSDDTCLPLRPOEMEVDSE 621
Db 511 CLTWYKQFDKETYEKFLNADDAEKTTSDAKKSKSKSFQCATADAVKPDFF----- 565

QY 622 DEKDPWLREKTTITQIEEFDVNEGKVMKLNWLVHMKH-----GFIADNQNMHA 672
Db 566 -----YKTLSTIPKFKDLQWQHEDAEELTHLLDQHEELISAIDGLTDNE----- 612

QY 673 CMLFVENYGQIKKIKNLCRNFLHLVSMH---DFNLISIMSIDKAVTKLEMOQKLEKGE 729
Db 613 -----IQMLOSINDEQL-KVFFIRNLSRYKAEF-----IKNASPRKLEIKY----- 656

QY 730 SASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQK 775
Db 657 --GVINDSTEE-----NGWHE-----VSGSSKRGKTK 683

```

## RESULT 7

```

US-08-487-203A-8
; Sequence 8, Application US/08487203A
; Patent No. 5683904
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: Maine
; COUNTRY: U.S.A.
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,203A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/005,002
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: MIT-5091A3Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-203A-8

```

```

Query Match 3.7%; Score 150; DB 1; Length 912;
Best Local Similarity 18.5%; Pred. No. 9.4e-05;
Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;

QY 65 MTDAAAREQESLKKIKQPKLSLTSSVSRRGNVSTPPRHSSGSLTPPVTPPTPSSSFRS 124
Db 3 MQDANKESYSVYK-----TSSPPPTPTNNQIPIYQAPLQWGYTQ 45

QY 125 STPTPEPTQI--YRFLRTRNLI---APFLHRT-----LTYMSHRN-----SRTNKRKTF 169
Db 46 APYLYPTQIYAYSPF---NMVNQNQPIY-HQSGSPHLPFPQNNINGSGSTNNNNKKKW 100

```

```

QY 170 KVDDMLSKVEMKMGQESHLSAHLQLTFTGFHKNDK-----PSPNS-----E 213
Db 101 HNGTNN--NGSSNGGANS--GSMYKSHYHNNHIPPMA--SPNSGNSGMMKQOT 159
QY 214 NQCN-----SVTLVLVVKVCHKRDVSCPIRQVPTGKKQVPLIPDLNQ 259
Db 160 NSNGNGSSATSPSYSSNSQYDLKPDVTKLN-----LKENSNLILPLF--INTT 213
QY 260 KFG-----NPSLAVSSNEFPEN-----SHMVKSYLLFVTPGRREFNGM 302
Db 214 EAEFAAASVQVELNWKALNLSSELSSESVSEKSAHHHTKSHSI-----PKNEE 264
QY 303 INGETN--ENIDVNEELPARRKENREDGKTFVAQMTVFNKRLQLLDGCEVAMQEMEE 361
Db 265 VKTETHEEEDAHKKPHASKDAHELKKTEVK-----EDAKQDRNE 307
QY 362 CPISKRAATWETILGKRLPPPTFSQGTLOFTLRGTETNDKSTAPIA-----KPLATR 417
Db 308 KVIQEQATVLPVDKKE--PEESVEEN-----TSKTSPPSPPAKSWASATSD 356
QY 418 NSESLHQENKPGSVKPTOTIAVKSLLTDDLOTRKEKDTPENROKLRIFYOFLYNNTRQ 477
Db 357 AIKSRQANKTVSGMWTKTPISGTTAGVSSTNMAATIGKSSPL-----LSK 405
QY 476 QTEARDLHCPWCTLNCRKLSLLKHLKCHSRFIFNVYHPKGRIDVSNICYDGSYA 537
Db 406 QPKKDKKYPVPPSTKGIPLGSA--LRMCFDPDFISVYLRNKOVENKIPVHSII-----458
QY 538 GNPQDIHQPGFAPSGFVKTPTITHLVCRP-----KRTKASM-----577
Db 459 --PRGIINRANICF-----MSVLQVLLYKPFIDVINLSTNTRNSVTSSTSKLLDA 510
QY 578 -----SEFLESDG--EVEQRTYSSGHNRLYFSDTCLPLRQPEMEVDSE 621
Db 511 CLTMVKQFDKETEYKFFLENADDAKTTESDAKSKSKSFQHCATADAVKPDF-----565
QY 622 DEKQEWLREKTIQIEBPSVNEKEVEMKWLHLVWKH-----GFIADQNMHA 672
Db 566 -----YKTLSTIPKFKDQWGHQDAEFLLDQJHEBLISAIDGLTNE-----612
QY 673 CMLFVNYGQKIIKGNLORNFMLHJVSMH-----DFNLISIMSIDKAVTKLREMOQKLEKGE 729
Db 613 ---IQNMLQSIINDQL--KVFFIRNLSRYGAEF-----IKNASPLKELIEKY----656
QY 730 SASPANEITEEQNTANGFSEINSKEALETDSVSGVSKQSKQK 775
Db 657 --GVINDDSTEE-----NGWHE-----VSGSKRGKRTK 683

```

## RESULT 8

```

US-08-118-101A-2
; Sequence 2, Application US/08118101A
; Patent No. 5620892
; GENERAL INFORMATION:
; APPLICANT: Kurtz, Stephen E.
; APPLICANT: Knickerbocker, Aron M.
; APPLICANT: McCullough, John R.
; TITLE OF INVENTION: A STRAIN OF SACCAROMYCES CEREVISIAE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/118,101A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-118-101A-2
;
; Query Match 3.6%; Score 146.5; DB 1; Length 1235;
; Best Local Similarity 20.1%; Pred. No. 0.00033;
; Matches 154; Conservative 90; Mismatches 227; Indels 295; Gaps 39;
;
QY 134 YRFLRPNLIAPFLHRTLTVMSHRNSRTNIRKTKFKVDDMLSKV-----EWMK 182
Db 148 FKNRRTKTILRELTARTMTKRTGTQRTSYPRKQAKTDDFOEKLFGEMVNRDEQDSVH 207
QY 183 GEQESHLSAHLQLTFTGF-----FHKNDKPSNSENENQNSVTLVLLVKVCHKKR 233
Db 208 SDQNSHDSIRDSNNNTNNGSSGLDDFVKEDETTDNGEYQENN-----252
QY 234 KQVSCPIRQVPTGKKQVPLIPDLNQTKGNFPLAVSNFEP--SPNSHMKVSYLLF---289
Db 253 -----SYSTGVSSNTVADESINOKPSSLRFPDEP 283
QY 290 -RVTPRGRREFNGINGETNENIDVNEELPARRKENREDGKTFVAQMTVFNKRLQL 348
Db 284 HSKQRPAR-----VPSEKFAKRSRD-----ISPADMYASIMWL 318
QY 349 DGEYEVAMQEMEECP-----SKRATWETILGKRLPPPTFSQGTPL 392
Db 319 QQKHE--ATADEGPPLVIGSPADGTRYKSNVNLKATG--INGNKIKRDKGNESNTD 374
QY 393 QTLRWGTNDKSTAPIAKPLATNSSESLHOF--NKPGSVKPTOT-----IAVKESL 443
Db 375 QNSV--SSEAN--STASVS-----DESSLHTNFGNKVPSLR--TNHRSNGSPIAITNA 423
QY 444 TTDLOTREKDTPENROKLRIFYOFLYNNTR-----QOTEARDDDLHCWCTLNCRKLYS- 499
Db 424 ETD-----KKHGPS-----IQFDITKPPRKISKVSTFDDLN--FKSSVLYRKKASK 468
QY 500 --LLKHLKLCCHSRFIFNVYHPKGRIDVSI--NECYDGSYAGNPODIHQPGFAPSRNGP 556
Db 469 KYLMKHF-----PKASRIIRQQIKRRLSTGSIENK-----SSNNV 502
QY 557 VKRTPITHLVCRPKRTKAMSE-----FLESEDEGEVQOORTYSSGHNRL 601
Db 503 SQRKPT-----DMDDDDDDDDGDNNGNEEYFADNESGD--EDERVQOQSEP---546
QY 602 YFHSOTCLPLRQPEME-----VSEDEKDPFWLREKTTITQIEEFSVNE 645
Db 547 --HSDSELKSHQOOQEKHQLOQNLHRMYKTKSFDDNRRAVPMERSRTI--DMAEAKDNE 603
QY 646 -----GEKEVMKLNHL-----VMKHG--FIADNMNHACMLFVENYQOK 683
Db 604 LARTPDDFQWYQNWKAHKKPFRKGNWKKIFEHGFYASDSDRNYPDN---SNTG--658
QY 684 IIKQNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOQKLEKESASPANEITEEQN 743
Db 659 -----NSILHYAE-----SILHEDGS-----HKNGBEASSDSNENIYSTNG 695
QY 744 GT-----ANGFSEINSKEA-----LETDSVSGVSKQSKQK 776
Db 696 GSDHGNLNPYNDDEGYGLHFDTDYDLDPRHDLKSGSGKYL 741

```

[illegible]

1590 IAASDNCPIQKE-TSADCVHPLEKILLTEEL-----HQXTNBEK 1632  
 168 TFKVDDMLSKVB-KMKQGEBSHLSAHLQITFTFFHNDKPSNS-ENEONSUTVLEVL 225  
 1633 LLHEKNELEQAQVELKCEVE-HLMKSMIE-----SKSSLESQHEKHDEQOQLLA 1681  
 226 VVCHCKKRDVSCPIRQVPTGKQVPLIPDLNQTGPGNFPPLAVSSNEFEPNSHMKVSY 285  
 1682 LK-----QQQVVTQEK-----ELQOHTHEHLTAEV 1707  
 286 SLFVTRPGRREFNGMETNENIDVNEELPARKNR-----EDGKFTV 333  
 1708 DHLKENIELGLNFKNEAQKTTKEOCLLNENKELEQSOHRLQCEIEELMKSLDKES--- 1764  
 334 AQMTVPDKNRRLQLLDGVEVAMQMECPISKKRATWETILDGKRPUPPTFFSQGTLO 393  
 1765 ALETKESEQVINLQENEMVMEELKNSOR-----TVI-----AERDOLQ 1808  
 394 FTLRWGTNDKSTAPIAKPLATRNSESLSHQENKPGSVKPTQTIAVKESLTTDLQTRKEK 453  
 1809 DDLRESSVMSIETQDLAKA-----QALQOQK-----VQELTSQISVLQEK 1852  
 454 DTPNENRQKLRIFYQFLYNNTRQOT-BARDDLHCPWCTLNCRKLYSLKHLKLC--HSR 510  
 1853 ISLLEN-----QMLYNVATVKETLSERDOLN-----OSKQHLFSEIETLSLKEKE 1899  
 511 FIFNVVYHPK-GARIDVINSCEYDGSYAGNPQDIHQPGFAGSRNGPVKPTPTH--IL 566  
 1900 FALGAEKDKADAAKTIDITE-----KISNIEEQ-----LLQATNLKLTVERESLI 1948  
 567 VCRPK-RTKASMSFLESFED-----GEVQOQTYSSGHNRLYFHSDDTCLPLAPQEME----- 617  
 1949 QCKEALANTEHLRETLKSKDIALKMEQERDEAA--NKVIALTEKSSLEEQINENVT 2006  
 618 -VDSDEKDPWL-----REKTIQIEEFS-----VNEGEKEVM-----KLWNLHVMK 660  
 2007 LKEGEKETYLPQPSQQSSQVEELRESLKYDLOLEAEKEISEATNEIKNL-TAK 2065  
 661 HGFIADNQMNHACMLFVNYGQKIKNLCLNFMHLVSMHDFNLISIMSIDKAVTKLRE 720  
 2066 ISSLEEEILQNASIL-----NEAVSERENLRHSQQLVSELEQLSLTLKSRDHAPA--- 2116  
 721 MQCKLEKESASPAEEITEQNGTANGFNSINSKEKALETDS 763  
 2117 -QSKREKDEAVNKI--ASLAEBIKILTKEMDBFRDSKESLQEQS 2157

RESULT 11  
 US-08-353-700-1  
 ; Sequence 1, Application US/08353700  
 ; Patent No. 559919  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YEN, TIMOTHY J.  
 ; APPLICANT: RATTNER, JEROME B.  
 ; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
 ; TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,  
 ; TITLE OF INVENTION: AND METHODS OF USE  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
 ; STREET: 1601 MARKET STREET, SUITE 720  
 ; CITY: PHILADELPHIA  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/353,700  
 ; FILING DATE: 09-DEC-1994

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-1  
 Query Match 3.3%; Score 135; DB 1; Length 3248;  
 Best Local Similarity 20.3%; Pred No. 0.019;  
 Matches 180; Conservative 134; Mismatches 339; Indels 232; Gaps 44;  
 QY 21 LHFTPLADLI-----EH--IEDNHIDTDPVLEKQELQOQTYV-ALSYNRMWTDAAARE 72  
 DB 1402 MHFAELQEKFLSLQSEHKILHQCMSKMSLQ-----TYVDSLKAENLVSLNLENF 1456  
 QY 73 QBSLKKIKPKLSLTLSVSSVRGNSVTPRHSSGSLTPVTPPTPSSSFRSSTTEPTQ 132  
 DB 1457 QGDLVKEMQLG-----EGLVPSLSSCVDPSSSLSS--LGDS 1494  
 QY 133 IYRFLRNLIAPIFHLRILTYSHRNRTNKRITFKVDDML-----SKVEMKGEQES 187  
 DB 1495 FYRAL-----LEQTDMSLLNLEGAVSANQCSVDEVFCSSLTQYVDSLKA--EN 1542  
 QY 188 HSLSAHLQITFTGFPHKNDK-----PSPNSENQNSVLEVL-----LVKVCHKRD 235  
 DB 1543 LVLSTNLNR-NFQGDVLVKEMQLGLEGLVPSLSSCVDPSSSLSSLGDSFYRALLEQTDG 1601  
 QY 236 VS-----CPTRQV-----PTGKKQVPLIP-----DLNQTGPNFPPLAV 269  
 DB 1602 MSLLNLEGVSNANQCSVDEVFCSSLOEBNLRKTPSAPAKGVEELESCELVYRQSLK 1661  
 QY 270 SNEFEPNSHMKVSYSLFVRTRPGRREFNGMETNEN-----IDVNEELP 318  
 DB 1662 LEEKVESQGIKMKNETIQELEQLLSERQELDLCKRQYLSENEQMOQLTSVTLEMSKLA 1721  
 QY 319 ARKRNREDGKTFVA--QMTVFDKNRRLQLLDGVEVAMO--EMECPISKKRATWETIL 375  
 DB 1722 AEKKQTEQLSLEVARLQGLDLSR--SLLGITEDAIQGNESCDISKEHSEIT-- 1778  
 QY 376 DGKRLP-----PFETFSQGTPLQTLRWGTETN-DKSTAPIAKPLAT 416  
 DB 1779 --ERTPKHDVHOICDKDAQOQDLNLDIEKITETGAVKPTGECSEQSPDTPNPPGDEKTK 1836  
 QY 417 RNSESLEHQ--ENKPGSVKPTQTIAYKESLTTDLQTRKEKDTENENRQKLRIFYQLYNN 474  
 DB 1837 GSSEICISLFSFGPNALVPMDFLGNQEDI-HNLQLR-VKETSNEMLRLLHVI----- 1886  
 QY 475 TRQOTEARDLHCPWCTLNCRKLYSLKHLKLCNHSRFFIFNYVYHPKGIARIDVSNCEYD- 533  
 DB 1887 -----EDRD-----RKVESLLNEMKELDSKL-----HLQEVOLMTKEACIEL 1924  
 QY 534 ----GSYAGNPQDI-HRQPGFAFSRNGPVKRTPIPHIL-----VCRPKTKASKE----- 579  
 DB 1925 EXIVGELKENSDDLSEKLEYFSCDHQELLQORVETSEGLNSDLEMHADKSSREDIGDNVAK 1984  
 QY 580 -----FLESEDEGEVQRTYSSG--HNRLYFHSDDTCLPLRPOEMVEDSED-----EKD 625  
 DB 1985 VNSDKERFLDVEN-ELSRIRSEKASIEHEALYLEAD-----LEVQTEKLCLEKD 2034  
 QY 626 PEWLREKTIQIEEFSVDNVEGEKENVKLNHLVHMVGHFIADNQMNHACMLFVNYGQKII 685

Db 2035 NE-NKQVIVCLEELSVVTSERNQLR-GELDTMSKKTALDQLS-----EKMKETQ 2085  
 Qy 666 KXNLCNFMHLVSMHD-----FNLISIMSID-----KAVTKLREMOQKLEKESASPA 734  
 Db 2086 ELESQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLSLEKDSQALS 2145  
 Qy 735 NEEITEEQNTANGSEINSKEKAL---ETDSVSGVSKQKQKL 776  
 Db 2146 TKCELENQ-----IAQLN-KEKELLVKESESQARLSSEYKEL 2183

## RESULT 12

PCT-US95-16216-1  
 ; Sequence 1, Application PC/TUS9516216  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yen, Timothy J.  
 ; APPLICANT: Ratner, Jerome B.  
 ; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
 ; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street Suite 720  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/16216  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/353,700  
 ; FILING DATE: 09-DEC-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Reed, Janet B.  
 ; REGISTRATION NUMBER: 36,252  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100  
 ; TELEFAX: (215) 563-4044  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3248 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; PCT-US95-16216-1

Query Match 3.3%; Score 135; DB 5; Length 3248;  
 Best Local Similarity 20.3%; Pred. No. 0.019;

Matches 180; Conservative 134; Mismatches 339; Indels 232; Gaps 44;

Qy 21 LHPFTLADLI-----EH---IENHDIDTPRVLEKQELQPTVY-ALSYINRMTDAARRE 72  
 Db 1402 MEFAELQKFLSLQSEHKTLHQHCOMSSKMBELQ-----TYVDSLKAENLVLSNLRNF 1456  
 Qy 73 QBSLKKIQPKLSLTLSSSSVSRGNVSTPPRHSSGSLTPPVTPPTITPSSFRSSTPTPTQ 132  
 Db 1457 QGDLVKEMQLG-----EGLVPSLSSSCVDPSSSLSS--LGDSS 1494  
 Qy 133 IYRFLTRNLIAPIFLHRTLTYMSHNSRNTNKRKTFKVDML-----SKYEKMKGEBS 187  
 Db 1495 FYRAL-----LEQGTGMSLNLNLEGAVSANQCSVDEVFCSSLOTIVDSLKA--EN 1542

Qy 188 HSLSAHLQLTFTGFPHKNDK-----PSPNSENBNQSVTLVL-----LVKVCHKKRKD 235  
 Db 1543 LVLSTNLR-NFQGLVKEMQLGEBGLVPSLSSSCVDPSSSLSLGDSSEFYRALLBOTGD 1601  
 Qy 236 VS-----CPIRQV-----PTGKQVPLIP-----DLNQTKPGNFPPLAV 269  
 Db 1602 MSLLSNLEGVVSANQCSVDEVFCSSLQENLTKETPSAPAKGVEELESCEVYRQSLEK 1661  
 Qy 270 SSNEFPSPNSHMYKSYSLFRVTRPPGRREFNGINGETNEN-----IDVNEELP 318  
 Db 1662 LEEMESQGMKNKBIQELLESSEBQELDLCKRKQVLSENEQWQKLTSTVLEMSKLA 1721  
 Qy 319 ARRRKRNREDGEKTFVA--QMTVPDKNRRLLQLLDGEYEVAHQ-EMERCPISKKEATWETIL 375  
 Db 1722 AEKQTEQLSLELEVAELQIQGLDLSR-SLLGIDTDAIQGRNESCDISKEHSET-- 1778  
 Qy 376 DGRKLP-----PPETFSQGTLOFTLRWTGETN-DKSTAIAPLAT 416  
 Db 1779 --ERTPKHDVHQICDKDAQDLNLDIEKITETGAVKPTGSCSGEPDNTNVEPPGDKTQ 1836  
 Qy 417 RNSESLEHQ--ENKPGSVKPTQTIAVKESLTTDLQTRKEKDTNENRQKLRIFYQFLYNN 474  
 Db 1837 GSSECSLSLSPGPNALVPMDFLGNQEDI-HNQLR-VKETSNNELLHVI----- 1886  
 Qy 475 TRQQTAEARDLHCPWCTLNCRKLYSLKHLKCHSRPFIYVYHPKGARIDVINECYD- 533  
 Db 1887 ----EDRD-----RKVESLINEMKELDskl-----HLOEVQLMTKEACIEL 1924  
 Qy 534 ----GSVAGNPQDI-HRQPGFAFSRGNPVKRTPIHIL-----VCRPKRTKASME-- 579  
 Db 1925 EKIVGELKENSIDLSEKLEYFSCDHQELLQRTVTSBGLNSDLEHMDKSRREIDGNVAK 1984  
 Qy 580 -----FLESDGEVEQOQRTYSSG--HNRLYFHSDDTCLPLRPQEMVEDSD-----EKD 625  
 Db 1985 VNSWKEKRFIDVEN-ELSRIRSEKASIEHEALYLEAD-----LEVQTEKLCLEK 2034  
 Qy 626 PEMLREXTITQIEFSDVNEGEKVKWKLWNLHVMKHGFIADNQMHACMLFVENYQKII 685  
 Db 2035 NE-NKQXVIVCLESEELSVVTSERNQLR-GELDTMSKKTALDQLS-----EKMKETQ 2085  
 Qy 686 KKNLCNFMHLVSMHD-----FNLISIMSID-----KAVTKLREMOQKLEKESASPA 734  
 Db 2086 ELESQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLSLEKDSQALS 2145  
 Qy 735 NEEITEEQNTANGSEINSKEKAL---ETDSVSGVSKQKQKL 776  
 Db 2146 TKCELENQ-----IAQLN-KEKELLVKESESQARLSSEYKEL 2183

## RESULT 13

US-08-328-254-6  
 ; Sequence 6, Application US/08328254  
 ; Patent No. 5710022

; GENERAL INFORMATION:

; APPLICANT: Zhu, Xueliang

; APPLICANT: Lee, Wen-Hwa

; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/328,254

; FILING DATE: 24-OCT-1994



QY 236 V--SCPIRQVTKKQVPLIPDL-----NOTKGFNPPSLAVSSNEFFPSNHHMVKSYLL- 288  
Db 443 IILSSQLQPTATGSRSSRLSFRHGHGSQTSILGSIKRKSIMRMGRPTSPFPGSSFSNMG 502  
QY 289 FRVTRPGRRREF-----NGMINGETNENIDVNE-----ELPARRKRNRRED 327  
Db 503 SRSTSPTEKGFASQHQHATGASVQSDLEDDIDPRANVLNVGNMLSVGEAPVESTSKEED 562  
QY 328 GE-----KTFVAQMTVFDKRRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPP 382  
Db 563 KDVPDPPIANAMAEELSSSMRRRQSTSDV-----DEAPVSLSKTSSSTRLNG----- 607  
QY 383 PETFSQOPTLOFTLRWTGETND-----KSTAPIAKPLATRNSESLHGOENKPGSVKPTQT 436  
Db 608 -----LGVHSRNTSASDIDGVPKKSTLG-APPAHTSAQMORMNSFAS----- 651  
QY 437 IAVKESLTTDLQTRK---EKDTPNENRQKLR 464  
Db 652 -----QTKQVFGERTENSARESLR 671

RESULT 15  
US-09-020-222-26  
; Sequence 26, Application US/09020222  
; Patent No. 6111073  
; GENERAL INFORMATION:  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage  
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPS)  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/020,222  
; FILING DATE: 06-Feb-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/798419  
; FILING DATE: 02/07/1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: P1066r1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-3216  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 907 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
US-09-020-222-26

Query Match 3.2%; Score 128.5; DB 3; Length 907;  
Best Local Similarity 19.2%; Pred. No. 0.0096;  
Matches 98; Conservative 74; Mismatches 174; Indels 165; Gaps 22;  
QY 18 GCGLHPFTLADLIEHEDNHDTPRVLEKQLQOPTVALSYINRFMTDAARREQESLK 77  
Db 262 GTGTTPPLPEPDFKENGKLYD---IDQLISKAPSY---PYSSRPSASASLASSPTR 315  
QY 78 KKIQPKLSLTLSSSVSRGNVTPPRHSGSLTPPTPTPSS----- 120

Db 316 SAFRPKSTETVSEV-----VSSPP-----TSPLHSPVKPVYNEQEQVTEVELSIPVP 364  
QY 121 SFRSSTPTPTQIYRFLRNLRIAPIF--LHRTLITVMSHRNSETNIRKRTFKVDDMLSKV 178  
Db 365 SIQEAESQKPEVLGTSSMRSPSVTPTFEVAARPLTSMVRSRH----- 407  
QY 179 EKKMGQESHSLSAHLQLTFTGFFHKNDKPSP---NSENEQNSVTLEVLVVKVCHKKRKD 235  
Db 408 ---NAETEVOAIPAATDI-----SPEVKEGKSENAIT-----KDND 442  
QY 236 V--SCPIRQVTKKQVPLIPDL-----NOTKGFNPPSLAVSSNEFFPSNHHMVKSYLL- 288  
Db 443 IILSSQLQPTATGSRSSRLSFRHGHGSQTSILGSIKRKSIMRMGRPTSPFPGSSFSNMG 502  
QY 289 FRVTRPGRRREF-----NGMINGETNENIDVNE-----ELPARRKRNRRED 327  
Db 503 SRSTSPTEKGFASQHQHATGASVQSDLEDDIDPRANVLNVGNMLSVGEAPVESTSKEED 562  
QY 328 GE-----KTFVAQMTVFDKRRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPP 382  
Db 563 KDVPDPPIANAMAEELSSSMRRRQSTSDV-----DEAPVSLSKTSSSTRLNG----- 607  
QY 383 PETFSQOPTLOFTLRWTGETND-----KSTAPIAKPLATRNSESLHGOENKPGSVKPTQT 436  
Db 608 -----LGVHSRNTSASDIDGVPKKSTLG-APPAHTSAQMORMNSFAS----- 651  
QY 437 IAVKESLTTDLQTRK---EKDTPNENRQKLR 464  
Db 652 -----QTKQVFGERTENSARESLR 671

Search completed: August 25, 2004, 17:58:57  
Job time : 39 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 25, 2004, 18:15:39 ; Search time 407.5 Seconds  
(without alignments)  
6114.120 Million cell updates/sec

Title: US-09-874-162A-7  
Perfect score: 7690  
Sequence: 1 cccgcgcgcgcgcgcag.....tacataaagtccaataatat 4409

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xl  
-Q/cgn2\_1/USPTO\_spool\_p/US09874162/runat\_25082004\_171843\_2162/app\_query.fasta\_1.4551  
-DB=A\_Geneseq\_29Jan04 -OFT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874162 @CGN1.1.642 @runat\_25082004\_171843\_2162 -NCFU=6 -ICFU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_29Jan04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3725	48.4	776	5	ABG30843
2	3413	44.4	803	3	ABG00066 Human jAZ
3	3237	42.1	739	5	ABG30842 Human joi
4	2043	26.6	388	4	AUI5578 Human nov
5	2043	26.6	388	6	ABU55047 Human nov
6	1497	19.5	289	4	AUI5958 Human nov
7	1497	19.5	289	6	ABU55027 Human nov
8	1453.5	18.9	292	4	AAU16416 Human nov
9	1453.5	18.9	292	6	ABU55485 Human nov
10	1127.5	14.7	955	4	ABB71635 Drosophil

11	940	12.2	175	4	AAU16402
12	940	12.2	175	6	ABU55471 Human nov
13	733	9.5	278	4	AAU16381 Human nov
14	733	9.5	278	6	ABU55450 Human nov
15	661	8.6	243	7	ADB64250 Human pro
16	661	8.6	243	7	ADC06802 Human pro
17	564	7.3	143	4	AAU15931 Human nov
18	564	7.3	143	6	ABU55000 Human nov
19	490	6.4	243	5	ABG30841 Human jux
20	457	5.9	102	3	ABG03182 Human sec
21	429	5.6	202	3	ABG43387 Human ORF
22	280	3.6	171	4	ABB11690 Human SFP
23	238.5	3.1	604	3	AAU53933 An Os-MPC
24	232	3.0	241	4	ABG26463 Novel hum
25	230	3.0	498	3	AAU54168 Arabidops
26	227.5	3.0	367	3	AAU54169 Arabidops
27	224	3.0	767	4	ABBS8240 Drosophil
28	224	2.9	611	3	AAU53932 A MPC1 pr
29	221	2.9	445	3	ABG00060 VRN2 poly
30	217.5	2.8	440	3	ABG00061 VRN2 poly
31	212	2.8	241	4	ABG26463 Novel hum
32	211	2.7	692	6	ABP70524 Histone d
33	211	2.7	813	3	ABG01674 FIS2 prot
34	188	2.4	2013	3	ABG18265 Plasmodiu
35	176.5	2.3	1142	4	ABG70713 S cerevis
36	176.5	2.3	1142	6	ABR53123 Protein s
37	176.5	2.3	1349	4	AAU70854 C albica
38	161.5	2.1	583	5	ABG93227 C. albica
39	161	2.1	1875	6	ABR53560 Protein s
40	159.5	2.1	339	3	AAU54170 Arabidops
41	159.5	2.1	912	2	ABG36731 Ubiquitin
42	157.5	2.0	1388	7	ADB80468 Ovarian c
43	156.5	2.0	3696	5	ABP40235 Scaphyloc
44	155	2.0	1087	2	AAU19935 B. burgdo
45	155	2.0	1119	2	AAU19934 B. burgdo

## ALIGNMENTS

RESULT 1

ABG30843

ID ABG30843 standard; protein; 776 AA.

AC ABG30843;

XX ABG30843;

DT 21-OCT-2002 (first entry)

DE Human jAZF1/jJAZ1 fusion protein.

XX Human jAZF1; juxtaposed with another zinc finger; jJAZ1; jAZF1/jJAZ1;

KW joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;

KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;

XX fusion protein.

XX Homo sapiens.

OS Synthetic.

OS Chimeric.

XX Key

Misc-difference 13..37

FT /note= "Encoded by ACTGCGCATCGGGGGCTGGGACTCCATTC-

FT CCACCGCTGGCGACCTCATCGACATCGAGCACACAC. C2H2 zinc

FT finger domain"

FT Region 99..100

FT /note= "Breakpoint for production of fusion protein

FT jAZF1/jJAZ1"

FT Misc-difference 357

FT /note= "Encoded by CAG"

FT Misc-difference 485..508

FT /note= "Encoded by CTGCAATGCGCTGGTGTACTCTGAATCGCGCAA-

FT ACTTTATAGTTTACTCAAGCATCTTAACCTGCCAT. C2H2 zinc finger

FT domain"

FT Region 558..575

Location/Qualifiers

/note= "Bipartite nuclear localisation signal"

FT XX  
PN XX  
XX XX  
PD 13-DEC-2001.  
XX XX  
PF 04-JUN-2001; 2001WO-US017936.  
XX XX  
PR 02-JUN-2000; 2000US-0209093P.  
XX XX  
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
XX XX  
PI Koontz J, Sklar J;  
XX XX  
DR WPI; 2002-575047/61.  
DR N-PSDB; ABK89163.  
XX XX  
PT Novel JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides useful as immunogens or  
PT antigens to raise or test anti-JAZF1, jJAZ1 or JAZF1/jJAZ1 antibodies.  
XX XX  
PS Claim 7; Fig 3; 76pp; English.  
XX XX  
CC The present invention relates to a new JAZF1 (juxtaposed with another  
CC zinc finger), jJAZ1 (joined with JAZF1) or JAZF1/jJAZ1 polypeptide. The  
CC methods of the invention can be used to identify a compound which  
CC controls proliferation of endometrial stroma, by expressing jJAZ1 in the  
CC presence of the compound, and determining whether the compound affects  
CC expression of jJAZ1, jJAZ1 or JAZF1/jJAZ1 polypeptides are useful  
CC as immunogens or antigens to raise or test anti-JAZF1, jJAZ1 or  
CC JAZF1/jJAZ1 antibodies. The invention can be used as bait proteins in a  
CC two hybrid assay or three hybrid assay to identify other proteins which  
CC bind or interact with JAZF1/jJAZ1-binding proteins. JAZF1, jJAZ1 or  
CC JAZF1/jJAZ1 molecules are useful for identifying the origin of tumour and  
CC as tumour marker protein to verify that a stromal tumour is from  
CC endometrium. The antibody is useful for promoting or decreasing fertility  
CC or pregnancy, and also for treating endometrial stromal tumours. The  
CC present amino acid sequence represents the human JAZF1/jJAZ1 fusion  
CC protein of the invention  
XX XX  
SQ Sequence 776 AA;

Alignment Scores:  
Pred. No.: 0 Length: 776  
Score: 3725.00 Matches: 726  
Percent Similarity: 93.56% Conservative: 0  
Best Local Similarity: 93.56% Mismatches: 50  
Query Match: 48.44% Indels: 0  
DB: 5 Gaps: 0

US-09-874-162A-7 (1-4409) x ABG30843 (1-776)

QY 52 ATGACAGCATCGCGCGCGCTCTCTCTCCCAATCCTGCCGATCGGGGGCTGCGGA 111  
Db 1 MetThrGlyLeuAlaAlaAlaSerPhePheSerAsn\*\*\*\*\* 20

QY 112 CTCACATTCCTCCACCTCGCGCGACCTCATCGAGCACATCGAGCAACACATCGATACA 171  
Db 21 \*\*\*\*\* 40

QY 172 GATCCACGGTATTAGAAAACAGAAATACAGCAGCCCACTATGTGCCCTCAGTTAC 231  
Db 41 AspProArgValLeuGluGlyGlnGluLeuGlnProThrValAlaLeuSerTyr 60

QY 232 ATAAATAGATTATGACAGATGCTGCCCGCGAGCAGCAGGAGTCCCTAAAGAAGAGATT 291  
Db 61 IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysile 80

QY 292 CAGCCGAGCTCTCGCTGACTCTGTCCAGCTCAGTGTCTCAGGGAATGTCTCCACTCCC 351  
Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100

QY 352 CCACGCCACAGAGTGGAGCGCTTACTCCCGCGTGCACCCACCATCACCCTCCCTCT 411

101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSerSer 120  
412 TCATTCCGCGAGCAGCACTCCGACAGAGCCAAACACAGATCTATAGATTTCTTCGAACCTGG 471  
121 SerPheArgSerSerThrProThrGluProThrGlnIleTyrArgPheLeuArgThrArg 140  
472 AATCTCATAGACCAATATTTTTCACAGAACTCTTACTACATGCTCTCATCAAACTCC 531  
141 AsnLeuIleAlaProIlePheLeuHisArgThrLeuThrTyrMetSerHisArgAsnSer 160  
532 AGAACAAACATCAAAAGCGAAACATTTAAAGTTGATGATATCTTATCAAAAGTAGAGAAA 591  
161 ArgThrAsnIleLysArgLysThrPheLysValAspAspMetLeuSerLysValGluLys 180  
592 ATGAAAGGAGAGCAAGAATCTCATAGCTTGTGAGCTCATTTTCAGCTTACGTTTACTGGT 651  
181 MetLysGlyGluGlnGluSerHisSerLeuSerAlaHisLeuGlnLeuThrPheThrGly 200  
652 TTCTTCCCAAAATGATAAGCCATCACCAACTCAGAAATGAACAAATTCGTTTACC 711  
201 PhePheHisLysAsnAspLysProSerProAsnSerGluAsnGluGlnAsnSerValThr 220  
712 CTGGAAGTCTCTCTTGTGAAAGTTTGCACAAAGAAAGAGATGTAAAGTTGTCCAATA 771  
221 LeuGluValLeuLeuValLysValCysHisLysLysArgLysAspValSerCysProIle 240  
772 AGCGAAGTCCACAGGTAAAGCAGGTGCTTTGATTCTCTGACCTCAATCAACAAACAAA 831  
241 ArgGlnValProThrGlyLysLysGlnValProLeuIleProAspLeuAsnGlnThrLys 260  
832 CCGGAAATTTCCGCTCTCTTCCAGTTTCCAGTAAATGAATTTGAACCTAGTAACAGCCAT 891  
261 ProGlyAsnPheProSerLeuAlaValSerAsnGluPheGluProSerAsnSerHis 280  
892 ATGTGGAAGTCTTACTGCTGCTGCTTATAGAGTACTGCTCCAGGAGAGAGAGTTAAT 951  
281 MetValLysSerTyrSerLeuLeuPheArgValThrArgProGlyArgArgGluPheAsn 300  
952 GGAATGATTAATGAGAAACCAATCAAAATATATGATGTCAATGAGAGCTTCCAGCCAGA 1011  
301 GlyMetIleAsnGlyGluThrAsnGluAsnIleAspValAsnGluLeuProAlaArg 320  
1012 AGAAAACGAAATCTGAGATGGGAAAGACATTTGTTGCAAAATACAGATATTTCAT 1071  
321 ArgLysArgAsnArgGluAspGlyGlyLysThrPheValAlaGlnMetThrValPheAsp 340  
1072 AAAACACAGCGCTTACAGCTTTTACATGGGAATATGAAGTAGCCATCAGCAAAATGAA 1131  
341 LysAsnArgArgLeuGlnLeuLeuAspGlyGlyLysThrGluValAlaMet\*\*GluMetGlu 360  
1132 GAATGTCCAATGAAGCAAGAAAGACCAATGGGAGACTATTCTTGTGGGAAGAGGTG 1191  
361 GluCysProIleSerLysLysArgAlaThrTrpGluThrIleLeuAspGlyLysArgLeu 380  
1192 CTTCATTCGAAACATTTCTCAGGAGCTAGCTTGCAGTTCACTCTTCGTGGACAGA 1251  
381 ProProPheGluThrPheSerGlnGlyProThrLeuGlnPheThrLeuArgTrpThrGly 400  
1252 GAGACCAATGATAAATCTACGGCTCTATTGCGCAACCTCTTGGCCACTAGAAATTCAGAG 1311  
401 GluThrAsnAspLysSerThrAlaProIleAlaLysProLeuAlaThrArgAsnSerGlu 420  
1312 AGTCTCCATCAGAAAACAGCGCTGCTCAGTTAAACCTACTCAAACTATTCGCTGTTAAA 1371  
421 SerLeuHisGlnGluAsnLysProGlySerValLysProThrGlnThrIleAlaValLys 440  
1372 GAATCAATTCACATACAGATCTACAAACAGAAAGAAAGAGTACTCCAAATGAAACCGA 1431  
441 GluSerLeuThrThrAspLeuGlnThrArgLysGluLysAspThrProAsnGluAsnArg 460  
1432 CAAAATTAAGAAATATTTTATCAGTTTCTCTATATAACAATACAGGCAACAACTGAA 1491  
461 GlnLysLeuArgIlePheTyrGlnPheLeuTyrAsnAsnAsnThrArgGlnGlnThrGlu 480



QY 856 GTTCCAGTAATGAATTTGAACCTAGTAACGCCATATGGTGAAGTCTTACTCGTGTGCTA 915  
 Db 296 ValSerAsnGluPheGluProSerAsnSerHisMetValLysSerTySerLeuLeu 315  
 QY 916 TTTAGAGTACTCGTCCAGGAAGAGAGAGAGTTTAAATGAATGATTAATGGAGAAACCAAT 975  
 Db 316 PheArgValThrArgProGlyArgArgGluPheAsnGlyMetIleAsnGlyGluThrAsn 335  
 QY 976 GAAATATGTGTCATGAAGAGCTCCAGCCAGAGAAACGAATCGTACGATGGG 1035  
 Db 336 GluAsnIleAspValAsnGluGluLeuProAlaArgArgLysArgAsnArgGluAspGly 355  
 QY 1036 GAAAAAGACATTTGTCACAAATGACAGTATTTGATAAAAACAGCGCTTACAGCTTTTA 1095  
 Db 356 GluLysThrPheValAlaGlnMetThrValPheAspLysAsnArgGluLeuLeu 375  
 QY 1096 GATGGGAATATGAAGTAGCCATGCGAAGAAATGGAGAAATCCAAATAGCAAGAAAGA 1155  
 Db 376 AspGlyGluTyrgluValAlaMetGlnGluMetGluGluCysProIleSerLysLysArg 395  
 QY 1156 GCAACATGGGACATTTCTGATGGAGAGGCTCCCTCATTCGAAACATTTTCTCAG 1215  
 Db 396 AlaThrTrpGluThrIleLeuAspGlyLysArgLeuProPheGluThrPheSerGln 415  
 QY 1216 GGACTAGCTGAGTTCACCTCTCGTTGGACAGAGAGACCAATGATAATCTACGGCT 1275  
 Db 416 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla 435  
 QY 1276 CCTATTGGCAACCTCTTGGCCACTAGAAATTCAGAGAGTCCATCAGGAAACCAAGCCT 1335  
 Db 436 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 455  
 QY 1336 GGTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCATTCAGTACTCAAA 1395  
 Db 456 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGln 475  
 QY 1396 ACAAGAAAGAAAGGATCTCCAAATGAACACCGACAAATTAAGAAATTTTATCAG 1455  
 Db 476 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrgln 495  
 QY 1456 TTTCTCTATAACAATPACAAAGCAACAACTGCAAGCAAGAGATGACCTGCAATGCGCT 1515  
 Db 496 PheLeuTyrglnAsnAsnThrArgGlnGlnThrGluAlaArgAspLeuHisCysPro 515  
 QY 1516 TGGTGACTCTGAAGTCCGCAAACTTTATGTTACTCAAGCATCTTAACTCTGCCAT 1575  
 Db 516 TrpCysThrLeuAsnCysArgLysLeuTyrglnSerLeuLeuLysHisLeuLysLeuCysHis 535  
 QY 1576 AGCAGATTATCTCAACTATGTTTATCATCAAAAGTGTAGATAGATGTTTCTATC 1635  
 Db 536 SerArgPheIlePheAsnTyrglnValHisProLysGlyAlaArgIleAspValSerIle 555  
 QY 1636 AATGAGTCTTATGAGTCTTATGCAAGAAATCTCAGGATATTCATCCCAACTGGA 1695  
 Db 556 AsnGluCysTyrglnAspGlySerTyrglnAsnProGlnAspIleHisArgGlnProGly 575  
 QY 1696 TTTGCTTTTACTCCAGCGACCAAGTAAAGACACCTATCACATATTTCTGTGTC 1755  
 Db 576 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 595  
 QY 1756 AGGCCAAACGAAACAAAGCAGCATGTCTGAATTTCTGAATCTGAAGATGGGAGTA 1815  
 Db 596 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 615  
 QY 1816 GAAACGAAAGCAATATAGTAGTGGCCACATGCTGTGTATTTCCATAGTATGACTGC 1875  
 Db 616 GluGlnGlnArgThrTyrglnSerGlyHisAsnArgLeuTyrglnPheHisSerAspThrCys 635  
 QY 1876 TTACTCTCCGTCACAAAGAAATGGAAGTAGATAGTGAAGTAAAGGATCTCGAATGG 1935  
 Db 636 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTip 655

QY 1936 CTAAGAGAAAAACCAATTACACAAATTAAGAGTGTTCGTGTTAAATGAAGAGAGAAA 1995  
 Db 656 LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys 675  
 QY 1996 GAAGTGATGAATCTCTGGAATCTCATGTGATGAAGCATGGGTTTATTCTGACAAATCAA 2055  
 Db 676 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 695  
 QY 2055 ATGAATCATGCTGTATGCTGTTGTAGAAAAATTAAGACAGAAAAATTAATTAAGAGAAT 2115  
 Db 696 MetAsnHisAlaCysMetLeuPheValGluAsnTyrglnLysIleIleLysLysAsn 715  
 QY 2116 TTATGTCGAAATCTCATGCTCTAGTCAGCATGCATCTTAAATCTTATTAGCAT 2175  
 Db 716 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 735  
 QY 2176 ATGTCAATGATAAAGCTGTACCAAGCTCCGTGAATGCAGCAAAATTAAGAAAGGG 2235  
 Db 736 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 755  
 QY 2236 GAATCTGCTTCCCTCCAAACCAAGAAATAAATGAAGAACAAATGGACACAGCAATGGA 2295  
 Db 756 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 775  
 QY 2296 TTTAGTGAATTAATCAAAAGAGAAAGCTTTGGAAACAGATAGTGTCTCAGGGTTTCA 2355  
 Db 776 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 795  
 QY 2356 AAACAGAGCAAAAAACAAAACTC 2379  
 Db 796 LysGlnSerLysLysGlnLysLeu 803  
 RESULT 3  
 ABG30842  
 ID ABG30842 standard; protein; 739 AA.  
 AC ABG30842;  
 XX 21-OCT-2002 (first entry)  
 DT Human joined with jAZF1 (jJAZ1) protein.  
 DE Human; jAZF1; juxtaposed with another zinc finger; jJAZ1; jJAZ1/jJAZ1;  
 XX joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;  
 KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;  
 KW chromosome 17.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Region 92..93  
 FT /note= "Breakpoint for production of fusion protein  
 FT jAZF1/jJAZ1"  
 FT Misc-difference 448..471  
 FT /note= "Encoded by CTGCATGCCCTTGGTGTACTCTGAACCTGCCGCAA-  
 FT ACTTATAGTTTACTCAAGCATCTAAACTCTGCCAT. C2H2 zinc finger  
 FT domain"  
 FT Region 521..538  
 FT /note= "Bipartite nuclear localisation signal"  
 XX  
 FN WO200193805-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 04-JUN-2001; 2001WO-US017936.  
 XX  
 PR 02-JUN-2000; 2000US-0209093P.  
 XX  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
 XX  
 PI Koontz J, Sklar J;  
 XX WPI; 2002-575047/61.



|||||  
Db 632 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysLeuAsn 651  
QY 2116 TTATGTGCAAACTTCATGCTTCATCTAGTCAGCATGCACTTAACTTATTAGCATA 2175  
Db 652 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 671  
QY 2176 ATGTCAATAGATAAAGCTTTACCAAGCTCCGTGAATGCAGCAAAATAGAAAAGGGG 2235  
Db 672 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 691  
QY 2236 GAATCTGTTCCCTGCAAAACAGAAATACTGAAGACAAAATGGACAGCAAAATGGA 2295  
Db 692 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 711  
QY 2296 TTTAGTGAATTAATCTCAAAAGAGAAAGCTTTGGAAACAGATAGTCTCAGGGGTTTCA 2355  
Db 712 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 731  
QY 2356 AAACAGACAAACAAACAAACTC 2379  
Db 732 LysGlnSerLysLysGlnLysLeu 739  
RESULT 4  
AAU15978  
ID AAU15978 standard; protein; 388 AA.  
XX AC AAU15978;  
XX DT 07-NOV-2001 (first entry)  
XX DE Human novel secreted protein, Seq ID 931.  
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; opthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX OS Homo sapiens.  
XX FN WO200155322-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001341.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217436P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241825P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.









Db 1 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla 20  
 QY 1276 CCTATTGCCAACTCTGCGCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAGCCT 1335  
 Db 21 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 40  
 QY 1336 GGTTCAGTTAAACCTACTCAAACTATTGCTGTATAAGAAATCATCTACAGATCTACAA 1395  
 Db 41 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGln 60  
 QY 1396 ACAAGAAAGAAAGAGTACTCCAAATGAAGAACCCACAAAATTAAGAAATATTATCAG 1455  
 Db 61 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGln 80  
 QY 1456 TTCTCTATAACAACTACAGGCAACAACTGAAGCAAGAGATGACCTGCTGCTCCCT 1515  
 Db 81 PheLeuTyrAsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysPro 100  
 QY 1516 TGGTGTACTGAACTGCGGCAACTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT 1575  
 Db 101 TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHis 120  
 QY 1576 AGCAGATTTATCTTCAACTATGTTTATCATCCAAAGGTGCTAGGATAGATGTTTCTATC 1635  
 Db 121 SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle 140  
 QY 1636 AATGAGTGTATGATGGTCTCTATCAGGAAATCTCAGGATATTTCATCGCAACCTGGA 1695  
 Db 141 AsnGluGlyTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly 160  
 QY 1696 TTTGCTTTTGTAGTCGCAACGGACCACTTAAGAGAACACCTATCACACATATTCTGTGTC 1755  
 Db 161 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 180  
 QY 1756 AGGCCAAACGAAACAAAGCAGCATGCTGAATTTCTTGAATCTGAAGATGGGGAAGTA 1815  
 Db 181 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 200  
 QY 1816 GAACAGCAAGAACATATAGTGTGGCCACAACTCTGTATTTCCATAGTATACCTGTC 1875  
 Db 201 GluGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisArgAspThrCys 220  
 QY 1876 TTACTCTCCGTCACAGAAATGAAGTATGATGATGATGATGATGATGATGATGATGATG 1935  
 Db 221 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGlyLysAspProGluTrp 240  
 QY 1936 CTAAGAGAAACCAACATTACACAAATTCGAAGTCTTCTGATGTTTAAATGAAGGAGAGAA 1995  
 Db 241 LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys 260  
 QY 1996 GAAGTGTGAACCTCTGCAATCTCCATGTCATGAGCATGGGTTTATTGCTGACAAATCAA 2055  
 Db 261 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 280  
 QY 2056 ATGAATCATGCTGATGCTGTTTCTAGAAAATTAAGCAAGAAATTAATTAAGAAAGAT 2115  
 Db 281 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsn 300  
 QY 2116 TTATGTCGAACTTCACTGCTTCACTGTCAGCATGCACTTAACTTATTTATGACATA 2175  
 Db 301 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 320  
 QY 2176 ATGTCAATAGATAAAGCTGTTTACCAAGCTCCGTGAATTCGACGCAAAATTAAGAAAGGG 2235  
 Db 321 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 340  
 QY 2236 GAATCTGCTCCCTGCCAACGAAAGAAATAACTGAAGAACAAAATGGGACAGCAAAATGGA 2295  
 Db 341 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 360  
 QY 2296 TTTAGTGAATTAATCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGTTTCA 2355

Db 361 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 380  
 QY 2356 AAACAGAGCAAAACAAACAAACAACTC 2379  
 Db 381 LysGlnSerLysLysGlnLysLeu 388  
 RESULT 6  
 AAU15958  
 ID AAU15958 standard; protein; 289 AA.  
 XX AAU15958;  
 AC AAU15958;  
 XX 07-NOV-2001 (first entry)  
 DT Human novel secreted protein, Seq ID 911.  
 DE Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
 KW preservative; antiproliferative.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX W0200155322-A2.  
 PN W0200155322-A2.  
 XX 02-AUG-2001.  
 PD 02-AUG-2001.  
 XX 17-JAN-2001; 2001WO-US0001341.  
 PF 31-JAN-2000; 2000US-0179065P.  
 XX 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184564P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 14-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 18-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226682P.  
 PR 23-AUG-2000; 2000US-0227182P.  
 PR 30-AUG-2000; 2000US-0227009P.  
 PR 01-SEP-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 06-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 08-SEP-2000; 2000US-0230439P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234224P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 27-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236803P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.

PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250191P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-488783/53.  
 DR N-PSDB; AAS25945.  
 XX  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 PS Claim 11; SEQ ID NO 911; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Antibodies to the proteins can also be used in  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunoassays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, and to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed

Alignment Scores:  
 Pred. No.: 1.9e-138 Length: 289  
 Score: 1497.00 Matches: 283  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.65% Mismatches: 0  
 Query Match: 19.47% Indels: 0  
 DB: 4 Gaps: 0

US-09-874-162A-7 (1-4409) x AAU15958 (1-289)



PT New polypeptides and nucleic acids, useful in gene therapy for treating,  
PT inhibiting or preventing e.g. neural, immune system, muscular,  
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
PT renal disorders.

PS Claim 11; SEQ ID NO 911; 402pp; English.

XX The invention relates to human novel polypeptides and their associated  
CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
CC therapy for treating, inhibiting or preventing neural disorders, immune  
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
CC (e.g. congenital heart defects, Epstein's anomaly and hypoplastic left  
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
CC ABUS5748 represent human novel polypeptides of the invention

XX SQ Sequence 289 AA;

Alignment Scores:  
Pred. No.: 1.9e-138 Length: 289  
Score: 1497.00 Matches: 283  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.65% Mismatches: 0  
Query Match: 19.47% Indels: 0  
DB: 6 Gaps: 0

US-09-874-162A-7 (1-4409) x ABUS5027 (1-289)

QY 1528 AACTGCGCAACTTTATAGTTTACTCAAGCATCTTAACATCTGCCATAGAGATTTATC 1587  
DB 6 AspCysArgLeuTyrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle 25  
QY 1588 TTCACACTGTTTATCATCCAAAGGTCCTAGGATAGATGTTCTCAATCAATGCTGTAT 1647  
DB 26 PheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIleAsnGluCysTyr 45  
QY 1648 GATGCTCTCTGACGAGAAATCTCAGGATATTCATCCCAACCTGGATTTGCTTTTGT 1707  
DB 46 AspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGlyPheAlaPheSer 65  
QY 1708 CGCACGACCGACGTTAAGAGACACCTATCACATATTTCTTGTCGAGCCCAACGA 1767  
DB 66 ArgAsnGlyProValTyrArgThrProIleThrHisIleLeuValCysArgProLysArg 85  
QY 1768 ACAAAGCAAGCATGCTGAAATTTCTGAATCTGAAGATGGGGAAGTAGAACAGCAAGA 1827  
DB 86 ThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluValGluGlnArg 105  
QY 1828 ACATATAGTAGTGGCCAAATCGTGTGTTATTTCCATAGTAGTACCTGCTTACCTCTCGT 1887  
DB 106 ThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerAspThrCysLeuProLeuArg 125  
QY 1888 CCACAAGAAATGGAAGTAGTAGTGAAGATCAAAAGGATCCTGAATGCTGAAGAGAAAA 1947  
DB 126 ProGlnGluMetGluValAspSerGluAspGluLysAspProGluTyrLeuArgGluLys 145  
QY 1948 ACCATTACAAATTAAGAGTTTCTGATTTAATGAAGAGAGAAAGTAGTGAATGAAA 2007  
DB 146 ThrIleThrGlnIleGluPheSerAspValAsnGluGlyGluLysGluValMetLys 165  
QY 2008 CTCGTGAAATCCATGTCATGAGCATGGTTTATTCGTGCAATCAATCAATCAATCGCC 2067  
DB 166 LeuTyrAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGlnMetAsnHisAla 185  
QY 2068 TGTATGCTGTTTGTAGAAAAATTTGGACAGAAAAATTAAGAGAAATTTATGTCGAAC 2127

DB 186 CysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsnLeuCysArgAsn 205  
QY 2128 TTCATGCTTTTCATCTAGTCTAGCATGATGACATCTTAACTTATTAGCATATATGTCAATAGAT 2187  
DB 206 PheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIleMetSerIleAsp 225  
QY 2188 AAAGCTGTTTACCAAGCTCGTGAATGCGCAAAATAGAAAAGGGGGAATCTGCTTCC 2247  
DB 226 LysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGlyGlnSerAlaSer 245  
QY 2248 CTTGCAAAACGAAATTAACCTCAAGAACAAATGGACAGCAAAATGGAATTTAGTCAAAAT 2307  
DB 246 ProAlaAsnGluGluIleThrGluGlnAsnGlyThrAlaAsnGlyPheSerGluIle 265  
QY 2308 AACTCAAAAGAGAAAGCTTTGGAACAGATAGTCTCAGGGGTTTCAAAACAGAGCAA 2367  
DB 266 AsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSerLysGlnSerLys 285  
QY 2368 AAACAAAAATCT 2379  
DB 286 LysGlnLysLeu 289

RESULT 8  
AAU16416  
ID AAU16416 standard; protein; 292 AA.  
XX AC AAU16416;  
XX DT 07-NOV-2001 (first entry)  
XX DE Human novel secreted protein, Seq ID 1369.  
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX OS Homo sapiens.  
XX PN WO200155322-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001341.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-MAR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.

PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225577P.	PR	08-NOV-2000;	2000US-0246603P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-022679P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226868P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249218P.
PR	08-SEP-2000;	2000US-0231142P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232081P.	PR	17-NOV-2000;	2000US-0249300P.
PR	12-SEP-2000;	2000US-0233568P.	PR	01-DEC-2000;	2000US-0250160P.
PR	14-SEP-2000;	2000US-0232397P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232400P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232401P.	PR	06-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251473P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251868P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251869P.
PR	21-SEP-2000;	2000US-0234274P.	PR	08-DEC-2000;	2000US-0251989P.
PR	25-SEP-2000;	2000US-0234397P.	PR	11-DEC-2000;	2000US-0251990P.
PR	25-SEP-2000;	2000US-0234398P.	PR	05-JAN-2001;	2001US-0254097P.
PR	26-SEP-2000;	2000US-0234398P.	XX		
PR	26-SEP-2000;	2000US-0234548P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235836P.	XX		
PR	27-SEP-2000;	2000US-0236327P.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327P.	PI		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	DR	WPI; 2001-488783/53.	
PR	29-SEP-2000;	2000US-0236369P.	DR	N-PSDB; AAS266403.	
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	02-OCT-2000;	2000US-0236802P.	PT	New nucleic acid molecules	

CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed

Alignment Scores:  
 Pred. No.: 3,836-134 Length: 292  
 Score: 1453.50 Matches: 280  
 Percent Similarity: 96.25% Conservative: 2  
 Best Local Similarity: 95.56% Mismatches: 9  
 Query Match: 18.90% Indels: 2  
 DB: 4 Gaps: 1

US-09-874-162A-7 (1-4409) x AAU16416 (1-292)

QY	748	AGAAAGGATGAAGTGTCCAAATAGGCAAGTCCACAGATGAAAAAGCAGGTCCCTTG	807
DB	1	LysLysAspValSerCysProIle***GlnValProThrGlyLys***Gln***ProLeu	20
QY	808	ATTCCTGACCTCAATCAAAACAAACCCCGAAATTCCTGCTCCCTGCGAGTTCCAGTAAT	867
DB	21	AsnProAspLeuAsnGlnThrLysProGly***PheProSerLeuAlaValSerSerAsn	40
QY	868	GAAATTTGACCTAGTAACAGCCATATGCTGAAGTCTTACTCGTTGCTATTAGAGTGACT	927
DB	41	GluPheGluProSer***SerHisMetValLysSerTyrSerLeuLeuPheArgValThr	60
QY	928	COTCCAGGAAGAAGAGAGTTTAATGGAATGATTAATGGAAGAAACCAATGAAATATTGAT	987
DB	61	ArgProGlyArgGluPheAsnGlyMetIleAsnGlyGluThrAsnGluAsnIleAsp	80
QY	988	GTCATGAAAGAGTCCAGCAGAGAAAGAAAGAAATCGTAGGATGGGAAAGACATTT	1047
DB	81	ValAsnGluGluLeuProAlaArgGlyLysArgAsnArgGluAspGlyGluLysThrPhe	100
QY	1048	GTTCGACAAATGACAGTATTTGATAAAACAGGGCTTACAGCTTTTAGATGGGGAATAT	1107
DB	101	ValAlaGlnMetThrValPheAspLysAsnArgGluGlnLeuLeuAspGlyGlyLys	120
QY	1108	GAAATGAGGATCGAGAAATGGAAGATGTCATTAAGCAAGAAAGAAAGCAATGGGAG	1167
DB	121	***ValAlaMetGlnGluMetGluGlyCysProIleSerLysLysArgAlaThrTrpGlu	140
QY	1168	ACTATCTTGTGGAGAGCTGCTCCATTCGAACATTTCTCAGGGACCTACCTG	1227
DB	141	ThrIleLeuAspGlyLysArgLeuProPheGluThrPheSerGlnGlyProThrLeu	160
QY	1228	CAGTTCACTCTTGTGGACAGGAGACCAATGATTAATCTACGGTCTCTATTGCCAAA	1287
DB	161	GlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAlaProIleAlaLys	180
QY	1288	CCTCTTGGCCTAGAAATCAGAGCTCTCCATCAGGAAACACACCCCTGGTTCAGTTAAA	1347
DB	181	ProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysProGlySerValLys	200
QY	1348	CCTACTCAAACTATTGCTGTTAAAGAAATCATTTGACTACAGATCTACAAACAAAGAAAGAA	1407
DB	201	ProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGlnThrArgLysGlu	220
QY	1408	AAGGATCTCCAAATGAAACCGCAAAATAGATATTTTATCAGTTCTCTATAAC	1467
DB	221	LysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGlnPheLeuTyrAsn	240
QY	1468	AACATACAGCAACAACTGAACAGAGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTG	1527
DB	241	AsnAsnThrArgGlnGlnThrGluAlaArgAspLeuHisCysProThrCysThrLeu	260
QY	1528	AACGCCGCAACTTTATGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTTATC	1587
DB	261	AsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle	280
QY	1588	TTCRACTATGTTATC---ATCCAAAGGCTGCTAGGA	1621

Db	281	PheAsn***Val-LeuSerSerGlnLysValLeuGly	292
RESULT	9		
ABU55485			
ID	ABU55485	standard; protein; 292 AA.	
XX	AC	ABU55485;	
XX	DT	18-MAR-2003 (first entry)	
XX	DE	Human novel polypeptide #572.	
XX	KW	Human; neural disorder; immune system disorder; renal disorder;	
XX	KW	muscular disorder; respiratory disease; reproductive disorder;	
XX	KW	gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;	
XX	KW	hyperproliferative disorder; inflammatory disease; allergic reaction;	
XX	KW	blood related disorder; cancer; immunosuppressive; antiinflammatory;	
XX	KW	cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;	
XX	KW	haemostatic; antiarteriosclerotic.	
OS		Homo sapiens.	
PN		US2002132753-A1.	
PD		19-SEP-2002.	
XX			
PF		17-JAN-2001; 2001US-00764864.	
XX			
XX	31-JAN-2000;	2000US-0179065P.	
PR	04-FEB-2000;	2000US-0180628P.	
PR	28-JUN-2000;	2000US-0214886P.	
PR	07-JUL-2000;	2000US-0216647P.	
PR	07-JUL-2000;	2000US-0216880P.	
PR	11-JUL-2000;	2000US-0217487P.	
PR	11-JUL-2000;	2000US-0217496P.	
PR	14-JUL-2000;	2000US-0218290P.	
PR	26-JUL-2000;	2000US-0220963P.	
PR	26-JUL-2000;	2000US-0220964P.	
PR	14-AUG-2000;	2000US-0224518P.	
PR	14-AUG-2000;	2000US-0224519P.	
PR	14-AUG-2000;	2000US-0225267P.	
PR	14-AUG-2000;	2000US-0225268P.	
PR	14-AUG-2000;	2000US-0225270P.	
PR	14-AUG-2000;	2000US-0225447P.	
PR	14-AUG-2000;	2000US-0225757P.	
PR	14-AUG-2000;	2000US-0225758P.	
PR	22-AUG-2000;	2000US-0226868P.	
PR	30-AUG-2000;	2000US-0228924P.	
PR	01-SEP-2000;	2000US-0229287P.	
PR	01-SEP-2000;	2000US-0229343P.	
PR	01-SEP-2000;	2000US-0229344P.	
PR	01-SEP-2000;	2000US-0229345P.	
PR	05-SEP-2000;	2000US-0229509P.	
PR	05-SEP-2000;	2000US-0229513P.	
PR	09-SEP-2000;	2000US-0231413P.	
PR	21-SEP-2000;	2000US-0234223P.	
PR	21-SEP-2000;	2000US-0234274P.	
PR	25-SEP-2000;	2000US-0234997P.	
PR	27-SEP-2000;	2000US-0235834P.	
PR	29-SEP-2000;	2000US-0236327P.	
PR	29-SEP-2000;	2000US-0236367P.	
PR	29-SEP-2000;	2000US-0236368P.	
PR	29-SEP-2000;	2000US-0236369P.	
PR	29-SEP-2000;	2000US-0236370P.	
PR	02-OCT-2000;	2000US-0236802P.	
PR	02-OCT-2000;	2000US-0237037P.	
PR	02-OCT-2000;	2000US-0237038P.	
PR	02-OCT-2000;	2000US-0237039P.	
PR	02-OCT-2000;	2000US-0237040P.	
PR	13-OCT-2000;	2000US-0239935P.	
PR	20-OCT-2000;	2000US-0240960P.	
PR	20-OCT-2000;	2000US-0241785P.	

20-OCT-2000; 2000US-0241809P.  
 01-NOV-2000; 2000US-0244617P.  
 17-NOV-2000; 2000US-0249299P.  
 08-DEC-2000; 2000US-0251856P.  
 08-DEC-2000; 2000US-0251868P.  
 08-DEC-2000; 2000US-0251869P.  
 (ROSE/) ROSEN C A.  
 (RUBE/) RUBEN S M.  
 (BARA/) BARASH S C.  
 Rosen CA, Ruben SM, Barash SC;  
 WPI; 2003-147444/14.  
 N-PSDB; ABX73744.  
 New polypeptides and nucleic acids, useful in gene therapy for treating,  
 inhibiting or preventing e.g. neural, immune system, muscular,  
 respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 renal disorders.  
 Claim 11; SEQ ID NO 1369; 402pp; English.  
 The invention relates to human novel polypeptides and their associated  
 polynucleotides. The polypeptides and polynucleotides are useful in gene  
 therapy for treating, inhibiting or preventing neural disorders, immune  
 system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 (e.g. congenital heart defects, Eisenstein's anomaly and hypoplastic left  
 heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
 ABUS5748 represent human novel polypeptides of the invention  
 SQ Sequence 292 AA;  
 Alignment Scores:  
 Pred. No.: 3.83e-134 Length: 292  
 Score: 1453.50 Matches: 280  
 Percent Similarity: 96.25% Conservative: 2  
 Best Local Similarity: 95.56% Mismatches: 9  
 Query Match: 18.90% Indels: 2  
 Gaps: 1  
 US-09-874-162a-7 (1-4409) x ABUS5485 (1-292)  
 QY 748 AGAAGGATGTAAGTGTCCATAGGCAAGTCCCAAGGTAAGCAAGGTCGCTTGT 807  
 Db 1 LysLysAspValSerCysProLeu\*\*\*GlnValProThrGlyLys\*\*\*Gln\*\*\*ProLeu 20  
 QY 808 ATTCTGACCTCAATCAACAAACAAACCCGGAATTTCCCGTCCCTGCAAGTTCCAGTAAT 867  
 Db 21 AsnProAspLeuAsnGlnThrLysProGly\*\*\*PheProSerLeuAlaValSerSerAsn 40  
 QY 868 GAATTTGAACCTAGTAACAGGCATATGCTGAGTCTTACTCGTTCCTATTAGAGTGAAT 927  
 Db 41 GluPheGluProSer\*\*\*SerHisMetValLysSerTyrSerLeuLeuPheArgValThr 60  
 QY 928 CGTCCAGCAAGAGAGCTTTAATGCAATTAATGAGAAACCAATGAATATTTGAT 987  
 Db 61 ArgProGlyArgArgGluPheAsnGlyMetIleAsnGlyGluThrAsnGluAsnIleAsp 80  
 QY 988 GTCAATGAGAGCTTCCAGCCAGAGAAACGAAATCGTGAGGATGGGAAAGACATTT 1047  
 Db 81 ValAsnGluGluLeuProAlaArgLysArgAsnArgGluAspGlyGluLysThrPhe 100  
 QY 1048 GTTGACAAATGACAGATTTTGATAAAACAGGCGCTTACAGCTTTAGATGGGAATAT 1107

Db 1101 ValAlaGlnMetThrValPheAspLysAsnArgArgLeuGlnLeuLeuAspGlyGluTyr 120  
 QY 1108 GAAGTACCATGACAGGAAATGGAAGATGTCCATATAGCAAGAAAGAGCAACATGGAG 1167  
 Db 121 \*\*\*ValAlaMetGlnGluMetGluGluCysProIleSerLysLysArgAlaThrTrpGlu 140  
 QY 1168 ACTATTCTTGATGGGAAGAGCGCTCCCTCCATTCGAAACATTTTCTCAGGAGACCTACGTTG 1227  
 Db 141 ThrIleLeuAspGlyLysArgLeuProPheGluThrPheSerGlnGlyProThrLeu 160  
 QY 1228 CAGTTCATCTTCTGTTGGACAGAGAGACCAATGATAAATCTACGGCTCCTATTGCCAAA 1287  
 Db 161 GlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAlaProIleAlaLys 180  
 QY 1288 CCTCTTCCCACTAGAAATTCAGAGAGCTCCATCAGCAAGAAACAAAGCCTGCTCAGTTAA 1347  
 Db 181 ProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysProGlySerValLys 200  
 QY 1348 CCTACTCAAACTATTGCTGTTAAAGAAATCAATGACTACAGATCTACAAACAAAGAAAGAA 1407  
 Db 201 ProThrGlnThrIleAlaValLysGluSerLeuThrAspLeuGlnThrArgLysGlu 220  
 QY 1408 AAGGTAATCTCAATGAAACCCGACAAATTAAGTAATTTTATCAGTTTCTCTATAAC 1467  
 Db 221 LysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGlnPheLeuTyrAsn 240  
 QY 1468 AACAAATCAAGGCAACAAACTGAAGCAAGAGATGACCTGCAATTCCTGCTACTCTG 1527  
 Db 241 AsnAsnThrArgGlnGlnThrGluAaArgAspAspLeuHisCysProTyrCysThrLeu 260  
 QY 1528 AACTGCGCGAACTTTATAGTTTACTCAAGCATCTTAACTGCCATAGCAGATTTATC 1587  
 Db 261 AsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuLysHisSerArgPheIle 280  
 QY 1588 TTCAACTATCTTTATC---ATCCAAAGGTGCTAGGA 1621  
 Db 281 PheAsn\*\*\*Val-LeuSerSerGlnLysValLeuGly 292  
 RESULT 10  
 ABB71635  
 ID ABB71635 standard; protein; 955 AA.  
 AC ABB71635;  
 DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 41697.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL15738.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX

PS Disclosure; SEQ ID NO 41697; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB16051), expressed DNA sequences (AB161840-AB16175) and the encoded proteins (AB161737-AB161742). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 955 AA;

Alignment Scores:  
Pred. No.: 1,23e-101 Length: 955  
Score: 1127.50 Matches: 273  
Percent Similarity: 49.46% Conservative: 137  
Best Local Similarity: 32.93% Mismatches: 240  
Query Match: 14.66% Indels: 179  
DB: 4 Gaps: 26

US-09-874-162A-7 (1-4409) x ABB71635 (1-955)

QY	244	ATGACAGATGCTGCCCGCAGAGACGAGTCCCTAAAGAGAGATTACGCGAGCTC	303
DB	1	MetAlaProAlaLysLysArgGluLysAspSer	11
QY	304	TCGCTGACTCTGCTCAGCTCAGTCTCTCGAGGGAAATGTG	351
DB	12	-----AsnProAspGlySerAlaAlaAsnGlyIleLeuGlyLeuThrHisGlyAlaPro	29
QY	352	CCAGCCACACAGTGGAGCTTACTCCCGCTGACCCACCCACCCCTCCTCT	411
DB	30	AspAlaSerAsnAlaGlySerThrValProThrAlaGluGlyGlnValLysLeuAsn	49
QY	412	TCATTCGCGCAGCAGCTCCG	453
DB	50	GlyHisGlnGlnGlnGlnGluPheLeuGlnAlaPheGluLysProThrGlnIleTyr	69
QY	454	AGATTCTTCGAATCTCGAATCTCATAGCACCATAATTTTTCACAGAACTTACTTAC	513
DB	70	ArgTyrLeuArgAsnArgHisGluThrAsnProIlePheLeuAsnArgThrLeuSerTyr	89
QY	514	ATGTCTCATCAAACTCCAGAACAAACATCAAAAGGAAACATTTAAAGTTGATGATG	573
DB	90	MetLysGluArgMetSerArgAsnAsnLysLysArgIleSerPheGlnValAsnSerMet	109
QY	574	TTATCAAAAGTAGAAATGAAAGGAGGACAGAACTCTATGCTTGTGCTGCTGCT	630
DB	110	LeuGluSerIle-----ThrGlnLysSerGluAlaValSerGlnAsnTyr	124
QY	631	TTGACAGCTTACGTTTACTGTTTCTCCACAAAATGATAGCCATCACCACAACTCAGAA	690
DB	125	LeuHisValIleTyrAspSerLeuHisGluLysLeuProAlaArgLeuAspAsnGluSer	144
QY	691	AATGAACAA-----AATTCTGTACCTCGAA	717
DB	145	GlyGluAspLeuLeuGlnGluLeuLeuLysGluAlaGlyGluSerValSerValGlu	164
QY	718	GTCCTGCTTGAAGTTGGCCACAAAAGAAAGGATGTAAGTTGCTCCAAATAGGCAA	777
DB	165	ThrThrLeuTyrLysIleThrArgSerLysArgLysAspSerThrLeuAspPheGlnGlu	184
QY	778	GTTCCACAGGTAAACAGCGCTTGTGATTCCTGACCTCAATCAACAAACCCCGGA	837
DB	185	LeuLeuSerLysCysSerGlnIleValTyrAsnPro-----LysAspArgValGly	201
QY	838	AATTTCCCGCTCCCTGCGAGTTTCCAGTAATGAATTTGAACT-----AGTAACAGCCATATG	894
DB	202	GluHisAlaThrIleSerIleProLeuGlnThrMetArgProMetGlyGluGlnHisThr	221

QY	895	GTGAAGTCTTACTCGTCTGCTATTTAGAGTACTCGTCCAGGAGAGAGAGAGTAAATGGA	954
DB	222	Leu-----TyrLysLeuLeuPheArgIleLys	230
QY	955	ATGATTAAATGAGAAACCAATGAAATAATTGATGTCAATGAGAGCTTCCAGCCAGAGA	1014
DB	231	ValLeuSerProSerThrCysAsn-----AspGluAsnAlaGluThrProPro-----Asn	247
QY	1015	AAACCAATTCGAGGATGGGAAAGACATTGTCACAAATGACAGTATTGTATAAA	1074
DB	248	LysArgSerArgProAsn-----GluLysMetPheGlySerGluLeuLeuLeuLysGluLys	266
QY	1075	AAACGGCGCTACAGCTTTTAGATGGGGAATATGAAGTAGCCATGACGAGAAATGAGAA	1134
DB	267	SerSerGlyPhe-----IleThrGluGlyGluTyrGluAlaMetLeuGlnProLeuAsnSer	285
QY	1135	TGTCCAATA-----AGCAAGAAAGACACATCGGAGACTATCTTGATGGGAG	1185
DB	286	ThrSerIleLysSerPheSerProLysLysCysThrTrpGluThrMetProAsp-----Ser	304
QY	1186	AGCTGCTCT-----CCATTGCAACATTTCTCAGGAGACCTACGTTGCAGTTCAC	1236
DB	305	TyrIleProLeuSerLeuThrTyrAspValTyrGlnGlnSerProMetLeuLysPheHis	324
QY	1237	CTTCGTTGGACAGGAGACCAATGATAATCTACGGCTCTATGCGCAACCTCTTGCC	1296
DB	325	LeuThrLeu-----SerAsnGluGlnLeuProGluMetIleSerAlaProGluLeu	341
QY	1297	ACTAGA-----AATTCAGAGAGCTC	1317
DB	342	GlnArgTyrValGlnHisLeuAspAlaValAlaGluMetAsnTyrAsnAsnAsnAsnTyr	361
QY	1318	CATCAGGAAAC-----AAGCCTGGTTCAGTTAAACCTTACTCAACTATT	1362
DB	362	AsnAsnAsnAsnAsnCysSerGlyLeuLysAsnGlySerGlyGlyGlyLysSerThrVal	381
QY	1363	GCTGTTAAAGTAATCAATTGACTACAGATCTACAAACAGAAAGAGATCTCCAAAT	1422
DB	382	Cys-----LysThrPro-----	386
QY	1423	GAACCCGACAAAATTAAGATAATTTTATCAGTTTCTTATAACAACAATACAGGCAA	1482
DB	387	-----GluHisIleGlnIleValTyrAsnPheMetTyrSerAsnAsnThrArgGln	403
QY	1483	CAAACTGAAGCAAGAGATGACCTGCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1542
DB	404	GlnThrGluTyrThrGlnGluLeuAsnCysProTyrCysGlyLeuAspCysLeuArgLeu	423
QY	1543	TATAGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTTATCTTCAACTGTTTAT	1602
DB	424	TyrAlaLeuLeuLysHisLeuLysLeuCysHisAlaArgPheAsnPheThrTyrGlnPro	443
QY	1603	CATCCAAAAGTCTAGGATAGATGTTTCTATCAATAGTGTATGATGCTGCTGCTGCTGCT	1662
DB	444	AlaGlySerGlyAlaArgIleAspValThrIleAsnAspAlaTyrAspGlySerTyrAla	463
QY	1663	GAATCTCAGATATTTCATCGCCAACTCGATTGCTTTTAGTCGCAAC-----GGACCA	1719
DB	464	GlySerProTyrAspLeuAlaGlyProSerGlySerSerPheAlaArgThrCysGlyPro	483
QY	1720	GTTAAGAGAACCTTACACATATTCTTGTGTGTCAGGCCCAACCAACCAACCAACCAAC	1779
DB	484	ValArgArgThrSerValThrSerLeuMetValCysArgProArgArgGlnLysThrCys	503
QY	1780	ATGCTGTAATTTCTGAATCTCAAGATCGGAGTAGTAGACAGCAACAGACATATAGTAG	1839
DB	504	LeuAspGluPheLeuLeuAspGluLeuAspGluLeuSerAsnGlnArgSerTyrIleThr	523
QY	1840	GCCCAATCTGCTGCTGATTTCCTATGATGATCTCTTACCTCTCTCCGTCACCAAGAAATG	1899
DB	524	GlyHisAsnArgLeuTyrHisHisThrGluThrCysLeuProValHisProLysGluLeu	543
QY	1900	GAAGTAGATAGTGAAGATGAAAGAGATCTGATGCTTAAAGAGAGAGAGAGATTTAATGGA	1959



[illegible]

PR	29-SEP-2000;	2000US-0236370P.	PT	New nucleic acid molecules encoding 461 human secreted proteins for
PR	02-OCT-2000;	2000US-0236802P.	PT	diagnosing, preventing, treating or ameliorating medical conditions and
PR	02-OCT-2000;	2000US-0237037P.	PT	used as food additives or preservatives.
PR	02-OCT-2000;	2000US-0237038P.	XX	Claim 11; SEQ ID NO 1355; 980pp; English.
PR	02-OCT-2000;	2000US-0237039P.	XX	The invention relates to isolated nucleic acid molecules and their
PR	02-OCT-2000;	2000US-0237040P.	XX	encoded secreted proteins. The nucleic acids and proteins are used to
PR	13-OCT-2000;	2000US-0239935P.	CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,
PR	13-OCT-2000;	2000US-0239937P.	CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
PR	20-OCT-2000;	2000US-0240960P.	CC	in diagnosing a pathological condition or susceptibility to a
PR	20-OCT-2000;	2000US-0241221P.	CC	pathological condition. Antibodies to the proteins can also be used in
PR	20-OCT-2000;	2000US-0241785P.	CC	alleviating symptoms associated with the disorders and in diagnostic
PR	20-OCT-2000;	2000US-0241786P.	CC	immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
PR	20-OCT-2000;	2000US-0241787P.	CC	(ELISA). Disorders which are diagnosed or treated include autoimmune
PR	20-OCT-2000;	2000US-0241809P.	CC	diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
PR	20-OCT-2000;	2000US-0241826P.	CC	neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
PR	01-NOV-2000;	2000US-0244617P.	CC	arrest, cerebrovascular disorders e.g. cerebral ischaemia, angogenesis,
PR	08-NOV-2000;	2000US-0246474P.	CC	nervous system disorders e.g. Alzheimer's disease, infectious caused by
PR	08-NOV-2000;	2000US-0246475P.	CC	bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
PR	08-NOV-2000;	2000US-0246476P.	CC	and many other disorders listed in the specification. The polypeptides
PR	08-NOV-2000;	2000US-0246477P.	CC	can also be used to aid wound healing and epithelial cell proliferation,
PR	08-NOV-2000;	2000US-0246478P.	CC	to prevent skin aging due to sunburn, to maintain organs before
PR	08-NOV-2000;	2000US-0246532P.	CC	transplantation, for supporting cell culture of primary tissues, to
PR	08-NOV-2000;	2000US-0246524P.	CC	regenerate tissues and in chemotaxis. The polypeptides can also be used
PR	08-NOV-2000;	2000US-0246525P.	CC	as a food additive or preservative to increase or decrease storage
PR	08-NOV-2000;	2000US-0246526P.	CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,
PR	08-NOV-2000;	2000US-0246527P.	CC	minerals, cofactors and other nutritional components. The present
PR	08-NOV-2000;	2000US-0246528P.	CC	sequence represents a novel secreted protein of the invention. Note: The
PR	08-NOV-2000;	2000US-0246532P.	CC	sequence data for this patent did not form part of the printed
PR	08-NOV-2000;	2000US-0246603P.	CC	
PR	08-NOV-2000;	2000US-0246610P.	CC	
PR	08-NOV-2000;	2000US-0246611P.	CC	
PR	08-NOV-2000;	2000US-0246613P.	CC	
PR	17-NOV-2000;	2000US-0249207P.	CC	
PR	17-NOV-2000;	2000US-0249208P.	CC	
PR	17-NOV-2000;	2000US-0249209P.	CC	
PR	17-NOV-2000;	2000US-0249210P.	CC	
PR	17-NOV-2000;	2000US-0249211P.	CC	
PR	17-NOV-2000;	2000US-0249212P.	CC	
PR	17-NOV-2000;	2000US-0249213P.	CC	
PR	17-NOV-2000;	2000US-0249214P.	CC	
PR	17-NOV-2000;	2000US-0249215P.	CC	
PR	17-NOV-2000;	2000US-0249216P.	CC	
PR	17-NOV-2000;	2000US-0249217P.	CC	
PR	17-NOV-2000;	2000US-0249218P.	CC	
PR	17-NOV-2000;	2000US-0249244P.	CC	
PR	17-NOV-2000;	2000US-0249245P.	CC	
PR	17-NOV-2000;	2000US-0249264P.	CC	
PR	17-NOV-2000;	2000US-0249265P.	CC	
PR	17-NOV-2000;	2000US-0249297P.	CC	
PR	17-NOV-2000;	2000US-0249299P.	CC	
PR	17-NOV-2000;	2000US-0249300P.	CC	
PR	01-DEC-2000;	2000US-0250160P.	CC	
PR	01-DEC-2000;	2000US-0250391P.	CC	
PR	05-DEC-2000;	2000US-0251030P.	CC	
PR	05-DEC-2000;	2000US-0251988P.	CC	
PR	05-DEC-2000;	2000US-0256719P.	CC	
PR	06-DEC-2000;	2000US-0251479P.	CC	
PR	08-DEC-2000;	2000US-0251856P.	CC	
PR	08-DEC-2000;	2000US-0251868P.	CC	
PR	08-DEC-2000;	2000US-0251869P.	CC	
PR	08-DEC-2000;	2000US-0251989P.	CC	
PR	08-DEC-2000;	2000US-0251990P.	CC	
PR	11-DEC-2000;	2000US-0254097P.	CC	
PR	05-JAN-2001;	2001US-0259678P.	CC	
XX	(HUMA-)	HUMAN GENOME SCI INC.	XX	
XX	Rosen CA,	Bazash SC, Ruben SM;	XX	
XX	WPI:	2001-488783/53.	XX	
XX	N-PSDB;	AAS26389.	XX	

QY	1447	TTTTATCAGTTTCTCTATAACAACAACTACAAGCGCAACAACTGACGAGATGACCTG	1506	
DB	1	PheTyrGln**LeuTyrAsnAsnThrArgGlnGlnThrGluAlaArgAspLeu	20	
QY	1507	CATTGCCCTTGCTGCTGACTCTGAACTGCCGCAACTTTATAGTTTACTCAAGCATCTTAA	1566	
DB	21	HisCysProTrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLys	40	
QY	1567	CTCTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGT	1626	
DB	41	LeuCysHisSerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAsp	60	
QY	1627	GTTTCTTCAATGAGTGTGTTATGATGCTCCCTATGAGGAATCTTCAGATATTCATGCG	1686	
DB	61	ValSerIleAsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArg	80	
QY	1687	CAACCTGGATTGCTTTTGTAGTCGCAACGACCGACAGTAAAGAGAACACCTTACACATATT	1746	
DB	81	GlnProGlyPheAlaPheSerArgGlnGlyProValLysArgThrProIleThrHisIle	100	
QY	1747	CTTGTGTGAGCCCAAAACGAAACAAAGCAACATGTCATGTTCTTCTTGAATCTGAAGAT	1806	
DB	101	LeuValCysArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAsp	120	
QY	1807	GGGAGTAGAACACAGACACATATAGTAGTGGCCCACTCGTCTCTATTTCATAGT	1866	
DB	121	GlyGluValGlnGlnArgThrTyrSerSerGlyHisAsnArg**TyrPheHisSer	140	
QY	1867	GATACCTGCTTACCTCTCCGTCACAAAGAAATGGAAGTAGATAGTGAAGATGAAGAT	1926	
DB	141	AspThrCysLeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAsp	160	

Alignment Scores:				
Pred. No.:	1.79e-83	Length:	175	
Score:	940.00	Matches:	172	
Percent Similarity:	97.73%	Conservative:	0	
Best Local Similarity:	97.73%	Mismatches:	3	
Query Match:	12.22%	Indels:	1	
DB:	4	Gaps:	0	

US-09-874-162A-7 (1-4409) x AAU16402 (1-175)

```
QY 1927 CCTGAATGCTAGAGAAACCAATACACAAATTGAGAGTTT 1972
Db ||||||| 161 ProGlutrrPleuArg-***LysProLeuHisLysLeuLysSerPhe 175

RESULT 12
ABU55471
ID ABU55471 standard; protein; 175 AA.
XX
AC ABU55471;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #558.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
XX US2002132753-A1.
PN
PX
PD
XX 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179063P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226758P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-147444/14.
DR N-PSDB; ABX73730.
XX
XX New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
PS Claim 11; SEQ ID NO 1355; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, Bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX
SQ Sequence 175 AA;

Alignment Scores:
Pred. No.: 1,79e-83 Length: 175
Score: 940.00 Matches: 172
Percent Similarity: 97.73% Conservative: 0
Best Local Similarity: 97.73% Mismatches: 3
Query Match: 12.22% Indels: 1
DB: 6

US-09-874-162A-7 (1-4409) x ABU55471 (1-175)
QY 1447 TTTTATCAGTTTCTCTATAACAACATACAAGSCAACAACTGAACGAGAGATGACCTG 1506
||| 1 PheTyrGln***LeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspLeu 20
QY 1507 CATGGCCCTGGTGTTACTCTGAACTCCGCAAACTTTATAGTTTACTCAAGCATCTTAAA 1566
||| 21 HisCysProTrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLys 40
QY 1567 CTCTGCCATAGCAGATTATCTTCAACTATGTTTATCTCATCCAAAGGCTGCTAGATAGAT 1626
||| 41 LeuCysHisSerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAsp 60
QY 1627 GTTTCCTATCAATGAGTGTATGATGCTCTCTATGCAGAGAAATCCTCAGATATTTTCATCG 1686
||| 61 ValSerIleAsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArg 80
QY 1687 CAACCTGGATTGCTTTTATAGTCGCAACGACAGTTAGAGAACCTATCACATATT 1746
||| 81 GlnProGlyPheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIle 100
QY 1747 CTTGTGTGCAGGCCAAACGAAACAAAGCAAGCATGCTCTGAATTTCTTGAATCTGAAGAT 1806
```



PR 08-NOV-2000; 2000US-0245613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-488783/53.  
 XX N-PSDB; AAS26368.  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 PS Claim 11; SEQ ID NO 1334; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Antibodies to the proteins can also be used in  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed

Alignment Scores:

Pred. No.: 6.76e-63 Length: 278  
 Score: 733.00 Matches: 142  
 Percent Similarity: 98.62% Conservative: 1  
 Best Local Similarity: 97.93% Mismatches: 2  
 Query Match: 9.53% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-874-162A-7 (1-4409) x AAU16381 (1-278)  
 QY 1 CCCGCCCGCGCTGCGAGAGCCGACACCCAGGGGGCTCTCGATGTAGCACCATGACAGGC 60  
 Db ProAlaAArgSerGlnAsnArgHisGlnGlySerGlyCysSerThrMetThrGly 38  
 QY 61 ATCGCCCGCGCTCTCTCTTCTCCATACCTGCGGATTCGGGGCTCGGACTCCACTTC 120  
 Db IleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe 58  
 QY 121 CCCACCTGGCCGACCTCATCGAGCAGCATCGAGGACAACACATCGATGATCCACGG 180  
 Db ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThrAspProArg 78  
 QY 181 GTTTAGAAAAACAAGAAATTACAGAGCCCAACCTATGTTCCCTGAGTTACATAATAGA 240  
 Db ValLeuGluLysGlnGluLeuGlnGlnProThrTyrValAlaLeuSerTyrlleAsnArg 98  
 QY 241 TTCATGACAGATGTCGCCCGCGAGAGCAGAGTCCCTAAAGAAAGATTTCAGCCGAAG 300  
 Db PheMetThrAspAlaAlaArgArgGluGlnGluSerLeuLysLysIleGlnProLys 118  
 QY 301 CTCGCTGACTCTGTCCAGCTCAGTGTCTCGAGGGAATGTGTCACATCCCCACGCCAC 360  
 Db LeuSerLeuThrLeuSerSerSerValSerArgGlyAsnValSerThrProProArgHis 138  
 QY 361 AGCAGTGGAGCCTTACTCCCGCGTGAACCCACCCATGACCCCTCTCTTCTTATCCGC 420  
 Db SerSerGlySerLeuThr\*\*\*ProValThrProIleThrProSerSerSerPheArg 158  
 QY 421 AGCAGCAGCTCCGACA 435  
 Db SerSerThrProThr 163  
 RESULT 14  
 ABUS5450  
 ID ABUS5450 standard; protein; 278 AA.  
 XX AC ABUS5450;  
 XX DT 18-MAR-2003 (first entry)  
 XX DE Human novel polypeptide #537.  
 XX KW Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX OS Homo sapiens.  
 XX PN US2002132753-A1.  
 XX PD 19-SEP-2002.  
 XX PF 17-JAN-2001; 2001US-00764864.  
 XX PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.

CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and  
 CC ABUS5748 represent human novel polypeptides of the invention  
 XX  
 SQ Sequence 278 AA;  
 Alignment Scores:  
 Pred. No.: 6.76e-63 Length: 278  
 Score: 733.00 Matches: 142  
 Percent Similarity: 98.62% Conservative: 1  
 Best Local Similarity: 97.93% Mismatches: 2  
 Query Match: 9.53% Indels: 0  
 DB: Gaps: 0  
 US-09-874-162A-7 (1-4409) x ABUS5450 (1-278)  
 QY 1 CCGCGCGCGCTCGCAGAGCGCAGCAGCGGGGGCTCTCGATGTAGCACCATGACAGGC 60  
 DB 19 ProAlaArgArgSerGlnAsnArgHisGlnGlyGlySerGlyCysSerThrMetThrGly 38  
 QY 61 ATCGCGCGCGCTCTCTTCTTCCAAATACCTGCCGATTCCGGGGCTCGGACTCCACTTC 120  
 DB 39 IleAlaAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe 58  
 QY 121 CCGACCTGGCGAGCTCATCGAGCAGATCGAGGACACACATCGATCAGATCCACGCG 180  
 DB 59 ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThrAspProArg 78  
 QY 181 GTTTAGAAAAACAAGAATTACAGCAGCAACCTATGTTCCTGAGTTACATAAATAGA 240  
 DB 79 ValLeuGluLysGlnGluLeuGlnGlnProThrTyrValAlaLeuSerTyrIleAsnArg 98  
 QY 241 TTCATGACAGATCTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 DB 99 PheMetThrAspAlaAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 118  
 QY 301 CTCCTGCTGACTCTGTCCAGCTCAGTGTCTCGAGGGAATGTCTCCACTCCCGCCAGCCAC 360  
 DB 119 LeuSerLeuThrLeuSerSerSerValSerArgGlyAsnValSerThrProProArgHis 138  
 QY 361 AGCAGTGAAGCTTACTCTCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 DB 139 SerSerGlySerLeuThr\*\*\*ProValThrProProIleThrProSerSerSerPheArg 158  
 QY 421 AGCAGCACTCCGACA 435  
 DB 159 SerSerThrProThr 163  
 RESULT 15  
 ADB64250  
 ID ADB64250 standard; protein; 243 AA.  
 XX  
 AC ADB64250;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human protein encoded by clone DFNES20076340.  
 XX  
 KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumor.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 PR  
 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0224526P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225575P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226686P.  
 PR 30-AUG-2000; 2000US-0226892P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251866P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR  
 (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 DR WPI; 2003-147444/14.  
 DR N-P5DB; ABX73709.  
 XX  
 New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 Claim 11; SEQ ID NO 1334; 402pp; English.  
 XX  
 The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,  
 CC congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory disorders (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood

XX (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

DR WPI; 2003-450961/43.  
DR N-PSDB; ADB62280.

PT New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.

PS Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a protein of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

Sequence 243 AA;

Alignment Scores:

Fragment No.:	8.32e-56	Length:	243
Pred. No.:	661.00	Matches:	128
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	8.60%	Gaps:	0
DB:	7		

US-09-874-162A-7 (1-4409) x ADB64250 (1-243)

QY	52	ATGACAGGCGATCGCGCGCGCTCTCTTCTTCCTCAATACCTCGCGATTCGGGGGCTCGGAA	111
Db	1	MetThrGlyIleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly	20
QY	112	CTCCACTCCCCACCCCTGGCGAGCTCATCGAGCACATCGAGAGACACCATCGATAC	171
Db	21	LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr	40
QY	172	GATCCACGGGTTTTGAAAAACAAGAATTACAGAGCCCAACCTATGTTGCCCTGAGTTAC	231
Db	41	AspProArgValLeuGluLysGlnLeuGlnGlnProThrTyrValAlaLeuSerTyr	60
QY	232	ATAAATAGATTCTACAGAGTCTCCCGCGAGAGAGAGTCCCTTAAGAGAGAGATT	291
Db	61	IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysIle	80
QY	292	CAGCGGAAGCTCTCGCTGACTGTGTCAGCTCAGTGTCTGAGGGAATGTGTCACCTCCC	351
Db	81	GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro	100





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 25, 2004, 18:26:55 ; Search time 78 Seconds  
(without alignments)  
5836.382 Million cell updates/sec

Title: US-09-874-162A-7

Perfect score: 7690  
Sequence: 1 ccgcgcggcgctgcagag.....tacataaagtccaataatat 4409

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp  
-Q/cgn2\_1/USFTO\_spool\_P/US09874162/runat 25082004 171845 2205/app query.fasta\_1.4551  
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874162 @CIGN 1 1 117 @runat 25082004 171845 2205 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRA=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238.5	3.1	604	US-09-339-947A-8	Sequence 8, Appli
2	224	2.9	611	US-09-339-947A-1	Sequence 1, Appli
3	159.5	2.1	912	US-07-789-915A-8	Sequence 8, Appli
4	159.5	2.1	912	US-08-005-002C-8	Sequence 8, Appli
5	159.5	2.1	912	US-08-487-203A-8	Sequence 8, Appli
6	159	2.1	1494	US-08-755-587-186	Sequence 186, App
7	156.5	2.0	3696	US-09-134-001C-5080	Sequence 5080, Ap
8	154.5	2.0	1388	US-09-572-191-2	Sequence 2, Appli
9	154.5	2.0	1388	US-09-723-262-2	Sequence 2, Appli
10	154.5	2.0	1388	US-09-723-219-2	Sequence 2, Appli
11	152.5	2.0	2482	US-08-328-254-6	Sequence 6, Appli
12	150.5	2.0	1167	US-09-803-671B-2	Sequence 2, Appli

13	146.5	1.9	1235	1	US-08-118-101A-2	Sequence 2, Appli
14	146	1.9	1972	4	US-09-418-710-21	Sequence 21, Appli
15	143.5	1.9	794	4	US-09-417-485D-8	Sequence 8, Appli
16	142.5	1.9	2954	4	US-09-150-867-1	Sequence 1, Appli
17	139	1.8	3248	1	US-08-353-700-1	Sequence 1, Appli
18	139	1.8	3248	5	PCT-US95-16216-1	Sequence 1, Appli
19	136	1.8	2184	4	US-09-417-485D-6	Sequence 6, Appli
20	134	1.7	1388	2	US-08-685-576-1	Sequence 1, Appli
21	134	1.7	2662	4	US-09-595-684B-31	Sequence 31, Appli
22	132.5	1.7	1969	4	US-09-418-710-72	Sequence 72, Appli
23	132.5	1.7	3418	2	US-08-639-501-2	Sequence 2, Appli
24	132.5	1.7	3418	2	US-08-603-753D-4	Sequence 4, Appli
25	132.5	1.7	3418	3	US-09-044-946-2	Sequence 2, Appli
26	132.5	1.7	3418	3	US-08-755-587-44	Sequence 44, Appli
27	132.5	1.7	3418	3	US-09-044-908-2	Sequence 2, Appli
28	132.5	1.7	3418	3	US-09-099-753-4	Sequence 4, Appli
29	132.5	1.7	3418	3	US-08-986-106-4	Sequence 4, Appli
30	131	1.7	1061	4	US-08-701-154A-5	Sequence 5, Appli
31	130.5	1.7	1036	3	US-08-891-640-3	Sequence 3, Appli
32	130.5	1.7	2329	3	US-08-755-587-16	Sequence 16, Appli
33	130	1.7	1032	4	US-09-914-259-26	Sequence 26, Appli
34	128.5	1.7	907	3	US-08-938-830-26	Sequence 26, Appli
35	128.5	1.7	907	3	US-09-020-222-26	Sequence 26, Appli
36	128.5	1.7	1031	4	US-09-914-259-24	Sequence 24, Appli
37	128	1.7	798	4	US-09-203-453-5	Sequence 5, Appli
38	128	1.7	798	4	US-09-900-236-5	Sequence 5, Appli
39	128	1.7	1164	4	US-09-457-708-2	Sequence 2, Appli
40	128	1.7	1164	4	US-09-950-046A-2	Sequence 2, Appli
41	128	1.7	1164	4	US-09-976-594-989	Sequence 989, App
42	128	1.7	1388	2	US-08-685-576-4	Sequence 4, Appli
43	128	1.7	2973	2	US-08-821-355A-7	Sequence 7, Appli
44	128	1.7	2973	2	US-09-003-687A-7	Sequence 7, Appli
45	128	1.7	2973	3	US-09-136-605-7	Sequence 7, Appli

## ALIGNMENTS

RESULT 1

US-09-339-947A-8  
; Sequence 8, Application US/09339947A  
; Patent No. 6630616  
; GENERAL INFORMATION:  
; APPLICANT: YOSHIDA, No. 6630616umasa  
; APPLICANT: YANAI, Yukihiko  
; APPLICANT: KATO, Yoshihiro  
; APPLICANT: HIRATSUKA, Junzo  
; APPLICANT: TAKAHASHI, Shigeru  
; APPLICANT: MIWA, Tatsushi  
; TITLE OF INVENTION: GENE FOR FLORAL REGULATION AND METHODS FOR CONTROLLING  
; FILE REFERENCE: 032735-001  
; CURRENT APPLICATION NUMBER: US/09/339,947A  
; CURRENT FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: JP 1998-180065  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: JP 1999-179043  
; PRIOR FILING DATE: 1999-06-24  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 604  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: ZN FING  
; LOCATION: (310)..(335)  
US-09-339-947A-8

Alignment Scores:  
Pred. No.: 4.7e-14 Length: 604  
Score: 238.50 Matches: 130  
Percent Similarity: 32.70% Conservative: 95  
Best Local Similarity: 18.90% Mismatches: 242

Query Match:	3.10%	Indels:	221
DB:	4	Gaps:	22
US-09-874-162A-7 (1-4409) x US-09-339-947A-8 (1-604)			
QY	436	GAGCAACACAGATCTATAGATTTCTTCGAACTCGGAATCTCATAGCACCACCAATATTTTG	495
DB	27	LySProValGluLeuTyrAsnIlelleGlnArgArgSerIleLysAsnProAlaPheLeu	46
QY	496	CACAGAACTCTTACTCATCTCTCAT-----CGAACTCCACAAACATC-----	543
DB	47	GlnArgCysLeuLeuTyrIleHisAlaArgArgIleLysArgSerLeuIleThIle	66
QY	544	-----AAAAGAAACATTTAAAGTTGATGATGTATTTATCAAAAGTAGAG	598
DB	67	SerLeuSerGlyGlyThrAsnLysGluLeuArgAlaGlnAsnIlePheProLeuTyrVal	86
QY	589	AAATGAAGAGAGCAAGATCTCATAGCTTGTCTCAGCTCAT-----	630
DB	87	LeuLeuAlaArgProThrAsnAsnValSerLeuGluGlyHisSerProIleTyrArgPhe	106
QY	631	-----TTGCAGCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCA	678
DB	107	SerArgAlaCysLeuLeuThrSerPheHisGluPheGlyAsnLysAspTyrThrGluAla	126
QY	679	-----CMAACTCAGAAAT-----GAACAAAATTTCTGTACCTCGGAAGTC	720
DB	127	ThrPheValIleProAspValLysAsnLeuAlaThrSerArgAlaCysSerLeuAsnIle	146
QY	721	CTGCTTGTGAAAGTTTGCACAAAAAGAAAG-----GATGAAGTTGTCCAATA	771
DB	147	IleLeuIleSer-----CysGlyArgAlaGluGlnThrPheAspAspAsnAsnCysSerGly	165
QY	772	AGGCAAGTTCCACAGSTPAAAAAGCAG-----GTG	801
DB	166	AsnHisValGluGlySerThrLeuGlnLysLeuGluGlyLysCysPheTrpGlyLysIle	185
QY	802	CTTTGATCTCGACCTCAATCAACAAACCCGGAATTTCCCGTCCCTCGCAGTTTCC	861
DB	186	Profile-----AspLeuLeuAlaSerSerLeuGlyAsnCysValSerLeuSerLeu---	202
QY	862	AGTAATGAATTTGAACCTAGTAACAGCCATATGTTGAAGTCTTACTCGTTGCTATTAGA	921
DB	202	-----	202
QY	922	GTGACTCTCCAGGAGAGAGAGATTAAATGAATGATTAATGAGAAACCAATGAAAT	981
DB	203	-----	208
QY	982	ATTGATGTCAATGAAGAGCTTCCAGCC-----AGAGAAACCAAAATCGTGAGATGGG	1035
DB	209	SerSerThrValGluMetThrProSerPheLeuGluProLysPheLeuGluAspSerSer	228
QY	1036	GAAAGACATTGTTGCAAAATGACAGTATTGATATAAACAGCGCTTACAGCTTTTA	1095
DB	229	CysLeuThrPheCysSerGlnLysValAspAlaThrGlySerPheGlnLeuGlnValSer	248
QY	1096	GATGGGAATATGAAGTAGCCATGACAGGAATGGAAGATCTCAATAAGCAAGAAAGA	1155
DB	249	IleSerAlaGlnGluAlaGlyAlaLysAspValSerGluSerProTyrSer-----	265
QY	1156	GCAACATGGGAGACTATTTGATGGGAAGGCTGCTCCATTCGAAACATTTTCTCAG	1215
DB	266	-----ValTyrSerTyrAsnAspValPro-----	274
QY	1216	GGACTACGTTGCAGTTTCACCTTCCTGTTGGACAGAGAGCAATGATTAATCTACGGCT	1275
DB	274	-----	274
QY	1276	CCTATTGCCAAACCTCTTCCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAAAGCCT	1335
DB	274	-----	274

[illegible]

Db 439 SerProAlaAspValGlnSerCysGlyAspProAspTyrValGlnSerIleAlaGlySer 458  
 QY 1777 AGCATGCTGAATTCCTTGAATCTGAAGATGGGAGTAGACAGCAA-----AGAACAA 1830  
 Db 459 ThrMetLeuGlnPheAlaLysThrArgLysIleSerIleGluArgSerAspLeuArgAsn 478  
 QY 1831 TATAGTAGTGGCCAAATCGTGTATTTCCATAGTAGTACCTCTTACCTCTCCGCTCA 1890  
 Db 479 ArgSerLeuLeuGlnLysArgGlnPhePheHisSerHisArgAlaGlnProMetAlaLeu 498  
 QY 1891 CAAGAAATG-----GAAGTAGATAGTAGAAGATGAAAGGATCTGAAATGGCTTAAGAA 1944  
 Db 499 GluGlnValLeuSerAspArgAspSerGluAspGluValAspAspValAlaAspPhe 518  
 QY 1945 AAAACCATACACAAATGAAGAGTTTCTGATGTTAATGAAGAGAGAGAGAGAGATG 2004  
 Db 519 GluAspArgArgMetLeuAspAspPheValAspValThrLysAspGluLysGlnMetMet 538  
 QY 2005 AAATCTCGAATCTCCATGTCATGAAGCATGGGTTTATTGCTGACAAATCAATGAATCAT 2064  
 Db 539 HisMetTrpAsnSerPheValArgLysGlnArgValLeuAlaAspGlyHisIleProTip 558  
 QY 2065 GCCTGATGCTGTTTGTAGAAAATATGACAGAGAAATTAAG---AAGAAATTTATGT 2121  
 Db 559 AlaCysGluAlaPheSerArgLeuHisGlyProIleMetValArgThrProHisLeuIle 578  
 QY 2122 CGAAATCTCATGCTTCATCTAGTCAGCATGCATGATCTTAACTTATTAGCATAATGTCA 2181  
 Db 579 TrpCysTrpArgValPheMetValLysLeuTrpAsnHisGlyLeuLeuAspAlaArgThr 598  
 QY 2182 ATAGATAAGCTGTTTACCAAGCTCGTGAATGACAG 2217  
 Db 599 MetAsnAsnAsnAsnThrPheLeuGluGlnLeuGln 610

RESULT 3

US-07-789-915A-8  
 ; Sequence 8, Application US/07789915A  
 ; Patent No. 5212058  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Rohan T.  
 ; APPLICANT: Tobias, John W.  
 ; APPLICANT: Varshavsky, Alexander  
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 19911108  
 ; APPLICATION NUMBER: US/07/789,915A  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brook, David E.  
 ; REGISTRATION NUMBER: 22,592  
 ; REFERENCE/DOCKET NUMBER: MIT-5091AA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 912 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein

US-07-789-915A-8  
 Alignment Scores:  
 Pred. No.: 2,278-06 Length: 912  
 Score: 159.50 Matches: 154  
 Percent Similarity: 33.14% Conservative: 130  
 Best Local Similarity: 17.97% Mismatches: 324  
 Query Match: 2.07% Indels: 249  
 DB: 1 Gaps: 31  
 US-09-874-162A-7 (1-4409) x US-07-789-915A-8 (1-912)  
 QY 20 GCCGACACACAGGGGGCTCTCGATGTACACATCGATACAGGATCGCCGCCCTCTCTCT 79  
 Db 6 AlaAsnLysGluGluSerTyr-SerMetTyrProLysThrSerSerProProProThr 25  
 QY 80 TCTCCAAATACCTGCCGATTCGGGGCTGGGACTCCACTTCCCACTCCGCGACCTCA 139  
 Db 26 -----ProThrAsn 28  
 QY 140 TCGAGCACATCGAGGACCAACACATCGATACAGATCCACGGGTTTGAAGAAACAAAGAA 199  
 Db 29 MetGlnIleProIleTyrGlnAlaProLeuGlnMetTyrGlyTyr--ThrGlnAlaProT 48  
 QY 200 TACAGAGCCAACTATGTTGCCGTAGTTACATAAATAGATTATCATGACAGATGCTGCC 259  
 Db 48 yrLeuTyrProThrGlnIleProAlaTyrSerPheAsnMetValAsnGlnAsnGlnProI 68  
 QY 260 GCCGAGAGCAGGAGTCCCTAAAGAAAGATTCACGCGAAGCTCTCGCTGACTCTGTCCA 319  
 Db 68 leTyrHisGlnSerGlySerProHisHisLeuProProGlnAsnAsnIleAsnGlyGlyS 88  
 QY 320 GTCAGTGTCTCGAGGAATGTGTCCACTCCCCACGCCACAGCAGTGGAGCCTTACTC 379  
 Db 88 erThrAsnAsnAsnAsnIleAsnLysLysLysLysLysLysLysLysLysLysLysLys 108  
 QY 380 CCCCCTGACCCCACTCCCT 439  
 Db 108 snAsnGlySerSerGlyAsnGlnGlyAlaAsnSerSerGlySerGlyMetSerTyrAsnL 128  
 QY 440 CAACACAGATCTATAGATTCTTCGAACCTCGGAATCTCATACACCAATATTTTGGACA 499  
 Db 128 ysSerHisThrTyrHisHisAsnTyrSerAsnAsnHisIle----- 141  
 QY 500 GAATCTTACTTACATGTCTCATCGAATCTCCAGAACAAACATCAAAAGGAAACATTTA 559  
 Db 142 -----ProMetAlaSerProAsnSerGlySerAsn----- 152  
 QY 560 AAGTTGATGATATGTTATCAAAAGTAGAGAAAATGAAAGAGAGAGCAAGAATCTCATAGCT 619  
 Db 153 -----AlaGlyMetLysLysGlnThrAsnSerSerAsnGlyAsnGlyS 167  
 QY 620 TGTAGCTCATTTGCGAGCTTAGTTTACGTTTCTTCCACAAAATATGATAGCCATCAC 679  
 Db 167 erSerAla-----ThrSerProSerT 174  
 QY 680 CAACTCAGAAAATGAACAAAATCTGTACCTCGAAGTCTGCTGTGAAAGTTTGGC 739  
 Db 174 yrSerSerTyrAsnSerSerSerGlnTyrAsp-----LeuTyrLysPheAspV 190  
 QY 740 ACAAAAAGAAAGGATGTAGTTTCCAAATAGGCAAGTCCCAAGGTAAAGAGCAGG 799  
 Db 190 alThrLysLeuLysAsn-----LeuLysGluAsnSerSerAsnLeuIleGlnL 206  
 QY 800 TGCCTTTGATTCCTGACCTCATCAACAAAACCCCGA----- 837  
 Db 206 euProLeuPhe-----IleAsnThrThrGluAlaGluPheAlaAlaAlaSerValGlnA 224  
 QY 838 -----AATTTCCCGTCCCTTGCGATTTCCAGTAAATTAATTTGAACCTAGTAAAC- 885  
 Db 224 rgTyrGluLeuAsnMetLysAlaLeuAsnLeuAsnSerGluSerLeuGluAsnSerSerV 244  
 QY 886 -----AGCCATATGGTGAAGTCTTACTCGTTGCTATTAGAGTGACTC 928

541	spAlaIysLysSerLysSerLysSerPheGlnHisCysAlaIthrAlaAspAlaValI	561
1886	GTCCACAGAAATGGAGTAGATAGTAGTGAAGATGAAAGGATCTCGAATGCTTAAGAGAAA	1945
561	ysProAspGluPhe-----TyrL--567	
1946	AAACCATTCACAAAATTGAAGAGTTTCTCATGCTTAATGAAGGAGAGAAAGAAGTGTATGA	2005
567	ysThrLeuSerThrIleProLysPheLysAspLeuGlnTrpGlyHisGlnGlnAspAlaG	587
2006	AACCTGTGGAAATCCCATGTCATGAAGCAT-----GGGT	2038
587	luGluPheLeuThrHisLeuLeuAspGlnLeuHisgluGluLeuIleSerAlaIleAspG	607
2039	TTATTGCTGCAATCAAAATCAATCATGCGCTGTATGCTTTTGTAGAAAAATATGGACAGA	2098
607	lyLeuThrAspAsnGlu-----IleGlnAsnMetLeuGlnS	619
2099	AAATAATAGAGAAATTTATGTGCGAAATTCATGCTTCATCTAGTCAGCATGCAT-----	2154
619	erIleAsnAspGluGlnLeu--LysValPhePheIleArgAsnLeuSerArgTyrGlyL	638
2155	----GACTTTTAATCTTATTAGCATAATGTCAATAGATAAAGCTGTTTACCAAGCTCGGTG	2209
638	ysAlaGluPhe-----IleLysAsnAlaSerProArgLeuLysG	651
2210	AAATGACGACAAAATTAGAAAAAGGGGAATCTGCTTCCCTCGCAACGAGAAATAACTG	2269
651	IuLeuIleGluLysTyr-----GlyValIleAsnAspAspSerThrG	665
2270	AAGACAAAATGGACAGCAAAATGGATTATTAGTCGAAATTAATCTCAAAAGACAGAGCTTTGG	2329
665	IuGlu-----AsnGlyTrpHisGlu-----	671
2330	AAACAGATAGTGTCTCAGGGGTTTCAAAACAGACAGCAAAAAACAAAAA	2376
672	-----ValSerGlySerSerLysArgGlyLysIleThrLys	683

US-06-005-002C-8  
 ; Sequence 8, Application US/08005002C  
 ; Patent No. 5494818  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Rohan T.  
 ; APPLICANT: Tobias, John W.  
 ; APPLICANT: Varshavsky, Alexander  
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kevin M. Farrell, P.C.  
 ; STREET: P.O. Box 999  
 ; CITY: York Harbor  
 ; STATE: Maine  
 ; COUNTRY: U.S.A.  
 ; ZIP: 03911  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/005,002C  
 ; FILING DATE: 15-JAN-1993  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/789,915  
 ; FILING DATE: 08-NOV-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Farrell, Kevin M.  
 ; REGISTRATION NUMBER: 35,505  
 ; REFERENCE/DOCKET NUMBER: MIT-5091AAZ  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 207-363-0558





QY	740	ACAAAAAAGAAAGGATGTAAAGTTGTCCAAATAAGGCAAGTTCCACAGGTAAAAAAGCAGG	799
Db	190	alThrLysLeuLysAsn-----LeuLysGluAsnSerSerAsnLeuIleGlnL	206
QY	800	TGCCTTTGATTCTGACCTCAATCAACAAACAAACCCGGA-----	837
Db	206	euProLeuPhe-----IleAsnThrThrGluAlaGluPheAlaAlaSerValGlnA	224
QY	838	-----AATTTCCTCCCTTCGAGTTTCAGTAATGAATTTGAACCTAGTAAC-	885
Db	224	rgTyrGluLeuAsnMetLysAlaLeuAsnLeuAsnSerGluSerLeuGluAsnSerS	244
QY	886	-----AGCATATGTTGAAGTCTTACTCTGTTCTATTAGAGTGACTC	928
Db	244	alGluLysSerSerAlaHisHisThrLysSerHisSerIle-----	258
QY	929	GTCCAGGAAGAAGAGATTTAATGGAATGATTAATGGAGAACCAAT---	985
Db	259	-----ProLysHisAsnGluValLysThrGluThrHisGlyGluGluGlnA	275
QY	986	ATGTCATGAAGACGTTCCAGCCAGAGAAACAAATCGTGAGGTGGGGAAGACAT	1045
Db	275	spAlaHisAspLysProHisAlaSerLysAspAlaHisGluLeuLysLysThrG	295
QY	1046	TTGTGTGCACAAATCACAGTATTTGATAAAAACAGCGCTTACAGCTTTTAGATGGGAAT	1105
Db	295	luValLysLys-----	298
QY	1106	ATGAGTAGCCATCCAGAAATGGAAGATGTCCAATAAGCAAGAAAGACACATGGG	1165
Db	299	--GluAspAlaLysGlnAspArgAsnGluLysValIleGlnGluProGlnAlaThrValL	318
QY	1166	AGACTATCTTGATGGGAAGGCTGCTCCATTTCGAAACATTTTTCACAGGACCTACGT	1225
Db	318	euProValValAspLysLysGlu-----ProGluGluSerValGluGluAsn-----	333
QY	1226	TGCAGTTCACTCTTCGTTGGACAGGAGACCAATGATAATCTACGGCTCTATTGCC-	1284
Db	334	-----ThrSerLysThrSerSerProSerProSerProSerProSerProAlaA	347
QY	1285	-----AAACCTCTTGCCACTAGAATTCAGAGAGTCTCCATCAGAGAAACAAAGC	1333
Db	347	laluysSerTrpSerAlaIleAlaSerAspAlaIleLysSerArgGlnAlaSerAsnLysT	367
QY	1334	CTGGTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCNTTGACTACAGATCTAC	1393
Db	367	hrValSerGlySerMetValThrLysThrProIleSerGlyThrThrAlaGlyValSerS	387
QY	1394	AAACAAGAAAAAGGATACTCCAAATGAAACCCGACAAAAATTAAGAATATTTTATC	1453
Db	387	erThrAsnMetAlaAlaAlaThrIleGlyLysSerSerProLeu-----	402
QY	1454	AGTTTCTCTATAACAACATACAGGCAACAACTGAAGCAAGAGATGACCTGCATTTGCC	1513
Db	403	-----LeuSerLysGlnProGlnLysLysAspLysLysTyrValP	416
QY	1514	CTTGGTGTACTGAACTGCGCAAACTTTATAGTTTACTCAAGCACTTTAAACTCTGCC	1573
Db	416	roProSerThrLysGlyLeuGluProLeuGlySerIleAla-----LeuArgMetCysP	434
QY	1574	ATAGCAGATTATCTCAACTAGTATTATCATCCAAAAGGTGTAGAGTAGATTTCTTA	1633
Db	434	heAspProAspPheIleSerTyrValLeuArgAsnLysAspValGluAsnLysIleProV	454
QY	1634	TCATAGTGTATGATGGCTCTCTATGCAGGAAATCTCTCAGGATATTTCATCGCCAACTG	1693
Db	454	alHisSerIleIle-----ProArgGlyIleIleAsnArgAlaA	467
QY	1694	GATTTGCTTTTAGTCGCAAGGACCACTTAAAGAGAACACCTATCACACATATTCTTGTT	1753
Db	467	snileCysPhe-----MetSerSerValLeuGlnValLeuLeuTyrC	481

Qy	1754	GCAGGCCA-----AAACGACAACAAAGCAAGCATG- 1782
Db	481	ylysProPheIleAspValIleAsnValLeuSerThrArgAsnThrAsnSerArgValG 501
Qy	1782	----- 1782
Db	501	lyThrSerSerCysLysLeuLeuAspAlaCysLeuThrMetTyrLysGlnPheAspLysG 521
Qy	1783	-----TCTGAATTCCTTGAACTCGAAGATGGG-----GAAGTAGAACACGACAAA 1825
Db	521	luThrTyrGluLysLysPheLeuGluAsnAlaaspAspAlaGluLysThrThrGlusErA 541
Qy	1826	GAACATATAGTAGTGCCCAACATCGTCTGTATTTCATAGTGNATACCTGTTACTCTCTCC 1885
Db	541	sPalalysLysSerLysSerLysSerPheGlnHisCysAlaThrAlaAspAlaValI 561
Qy	1886	GTCACAGAAGAAATGGAAGTAGATAGTCGAAGTCAAAGGATCCTGTAATGGCTPAAGAGAA 1945
Db	561	ysProaspGluPhe-----Tyrl 567
Qy	1946	AAACCATTACACAAATGAAGAGTTTTCTGATGTTAATGAAGGAGAGAAAGAAGTGATGA 2005
Db	567	ysThrLeuSerThrIleProLysPheLysaspLeuGlnTrpGlyHISgInGluAspAlag 587
Qy	2006	AACCTGGAAATCTCCATGTCATGAAGCAT-----GGGT 2038
Db	587	luguLuhelEuThrHisLeuLeuAspGlnLeuHisgluLuuIleSerAlaIleAspG 607
Qy	2039	TTATTCTGCACATCAATGAATCATGCGCTGTATGCTGTTGTAGAAAAATTATGGACAGA 2098
Db	607	lyLeuthraspAsnGlu-----IleGlnAsnMetLeuGlns 619
Qy	2099	AAATAATTAAAGAAAATTTATGTGGAACATTCATGCTTCATCTAGTCAGCATGCAT---- 2154
Db	619	erlleasnaspGluInLeu--LysvalPhePheileargAsnLeuSerArGtyrGlyL 638
Qy	2155	-----GACTTTAATCTTATTAGCATTAATGCAATAGATAAAGCTGTTACAGACTCCGGT 2209
Db	638	ysalaGluPhe-----IleLysAsnAlaserProArgLeuLysG 651
Qy	2210	AAATGACGACAAARAATTAGAAAGGGGAATCTGCTCCCTCGACACGAGAAATAACTG 2269
Db	651	luLeuIleGluLysTyr-----GlyValIleAsnAspAspSerThrg 665
Qy	2270	AAGAACAATAATGGGACAGCAAAATGGATTTAGTGAAATTAAC TCAAAAGAGAAAGCTTTGG 2329
Db	665	lugu-----AsnGlyTriPhisglu----- 671
Qy	2330	AAACAGATAGTCTCAGGGGTTTCAAAACAGACGACAAAAACAAAAA 2376
Db	672	-----ValSerGlySerSerLysArgGlyLysLysThrLys 683

RESULT 6  
US-08-755-587-186  
; Sequence 186, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; identification and sequencing of the BRCA2 cancer  
; TITLE OF INVENTION: susceptibility gene and uses thereof.  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Bell Seltzer Park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 25-NOV-1996
; PRIORITY DATE: 25-NOV-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-755-587-186

```

```

Alignment Scores:
Pred. No.: 3,21e-06 Length: 1494
Score: 159.00 Matches: 177
Percent Similarity: 33.92% Conservative: 151
Best Local Similarity: 18.30% Mismatches: 355
Query Match: 2.07% Indels: 284
DB: 3 Gaps: 46

```

US-09-874-162A-7 (1-4409) x US-08-755-587-186 (1-1494)

```

QY 118 TTCGCCACCTGGCGACCTCATCGAGCAGCATCGAGCAACCAACATCGATACAGATCCA 177
Db 370 PheLysThrAlaSerAsnLysGluLeuLysSerGluAsnLysLeuLysSerGluLeuVal 389
QY 178 CGGGTTTGTAGAAACACAGAAATACAGAGCAACCTATGTTGCC--CTGAGTTACATA 234
Db 390 MetLeuPheLysAspLeuGluHisTyrProThrAsnLeuAlaCysLeuGluLeuVal 409
QY 235 AATAGATTTCATGACAGATGCTCCCGCGAGAGAGAGTCCCTTAAGAGAAAGATTCCAG 294
Db 410 AsnThrSerLeuGluSerGlnLysLysProSerLysSer-----HisAlaLeuAsp 427
QY 295 CGAAGCTCTCCGACTCTGTCAGCTCAGTGTCTCGAGGAATGTCTCCACTCCCCCA 354
Db 428 ProGlnLeuAsnLeu-----IleSerGlyPheValGln-----AsnSerThrTyrValSer 444
QY 355 CGCCACAGCAGTGGAGCCCTTACTCCCGCGTGACCCACCCATCACCCTCTCTTCA 414
Db 445 AspSerGluSerGlyHisThrAlaProProThrLeuSerLeuLysGlnAspSer 464
QY 415 TTCGCGACAGCATCTCCGACAGAGCAACACAGATC----- 450
Db 465 AsnArgAsnLeuThrProSerGlnLysAlaGluLeuThrGluLeuSerThrIleLeuGlu 484
QY 451 -----TATAGATTCTCGAACTCCGAATCTCATAGCACCACATATTTTG 495
Db 485 GluSerGlySerGlnPheGluPheThrGlnPheArgLys-----ProSerHisIle 501
QY 496 CACAGAACTCTTACTTACATGCTCTCATCGAACTCCAGAAC-----AACATCAAAAGG 549
Db 502 IleGlnLysAsnProPheGluMetProGluAsnGlnLeuThrIleLeuAsnSerThrSer 521
QY 550 AAAACATTAAAGTTGATGATG-----TTATCAAAAGTAGAG 588
Db 522 LysGluTrpLysAspAspLeuHisLeuThrThrAsnAlaProSerIleGlnValAsp 541
QY 589 AAAATGAAA-----GGAGAGCAAGAATCTCATAGCTTGTGAGCT--- 627

```

```

Db 542 SerLysLysSerGluGlyIleIleGlyGlyLysGlnLysPheAlaCysLeuSerArgThr 561
QY 628 -----CAITTCGACGCTTACGTTTACTTGGT 651
Db 562 SerCysAsnArgSerAlaSerGlyTyrSerThrAspLysAsnGluValGluPheArgGly 581
QY 652 TTCTTCCAC---AAAAATGATAGCCATCACCAAACTCAGAA----- 690
Db 582 PheTyrSerAlaArgGlyThrLysLeuAsnValGlySerGluAlaLeuGlnLysAlaLys 601
QY 691 -----AATGAACAAAATTCCTGTT----- 708
Db 602 LysLeuPheSerAspLeuGluAsnIleAsnGluGluThrSerValGluValAspArgSer 621
QY 709 -----ACCTGGAAGTCTGCTTGTGTGAAAGTTTCCACAAA 744
Db 622 PheSerSerLysTyrAsnAspSerValSerMetIleGlnIleGluAspCysAsnAsp 641
QY 745 AAAAGAAAGGATGTAACTGTCCTCAATAAGCAAGTCCACACAGTAAAGAGAGGTCCT 804
Db 642 LysAsnAsnGlu-----ProAsnAsnLysCysArgLeu--- 652
QY 805 TTGATTCCTGACCTCAATCAAAACAAACCCGGAATTTCCGCTCCTTGCAGTTTCCAGT 864
Db 653 -----IleLeuGln 655
QY 865 AATGAATTTGAACTAGTACACGCCATATGCTGAAGTCTTACTCGTGTCTATTAGA--- 921
Db 656 AsnAsnIleGluMetThrThrAspIlePheValGluGluTyrThrGluSerTyrArgArg 675
QY 922 -----GTGACTCGTCCAGGAAGAGAGAGATTTTAATGGAATG 957
Db 676 AsnThrGluAsnGluGlyAsnGlnCysThrAspAlaGlyArgAsnThrCysAsnSerGlu 695
QY 958 ATTAATGGA-----GAAACCAATGAAATATTGATGTCATGAAGAG----- 999
Db 696 SerAspGlySerAspSerLysAsnAspThrValTyrIleHisGluGluAsnGly 715
QY 1000 CTCCACGCCAGAAAGAAACGAAATCGTAGAGTGGGAAAGACATTTGTTGCACAAATG 1059
Db 716 LeuProCysIleAspGlnHisAsnIleAsp-----LeuLysLeuPhe 729
QY 1060 ACAGTATTTCATAAACACGCGCTTACAGCTTTTATAGTGGGAATATGAA----- 1110
Db 730 SerGlnPheMetLysGluGlyAsnThrGlnIleLysGluGlyLeuSerAspLeuThrCys 749
QY 1111 GTAGCCATGACGAAATGGAAGATGTCCTCAATGACGAAGAAAGACCAACATGGAGACT 1170
Db 750 LeuValMetLysAlaGluGluThrSerHisValThrMetSerAsnLysGlnGlnLeuThr 769
QY 1171 ATTCTTGATGGAGAGCGCTGCTCCTCATTCGAAACATTT-----TCTCAGGACCTACG 1224
Db 770 AlaAsnThrGlyGlnAsnIleLysAspPheAspThrPheTyrLeuSerPheGlnThrAla 789
QY 1225 TTGAGTTCCTCTGTTGGACAGGAGACCAATGATTAATCTACGGCTCTCTATTGCC 1284
Db 790 SerArgLysAsnIleArgValSerArgLysLeuAsnLysAlaArgSerLeuLeuAsn 809
QY 1285 AAACCTCTTCCCACTAGAAAT-----TCAGAGAGTCTCCATCAGGAAACCAAG 1332
Db 810 GlnLysTrpThrGluGluGluLeuAsnAsnPheSerAspSerLeuAsnSerGluLeuLeu 829
QY 1333 CTGTGT---TCAGTTAAACCTACTCAAACTATTGCTGTTTAAAGATCATATTGACTACAGAT 1389
Db 830 ProGlyIleAspIleLysLysThrAsp-----IleSerAsnHisGluValIleGlu 846
QY 1390 CTACAAACAGAAAGAAAGAGTACT----- 1416
Db 847 AsnThrGluArgLysAspLysIleThrLysGluSerLeuIleGlyThrGluAsnIleLeu 866
QY 1417 -----CCAAATGAAACCCGCAAAATTAAGAATATTATTTATCAGTTT 1458

```

Db 867 LeuIleLeuGlnArgProGluSerLysLysIleLysLysIleLysGluSerAlaValLeu 886  
QY 1459 CTCATAACAACAAATACAGGCAACAACTGAAGCAAGAGATGACCTGCTGCTGG 1518  
Db 887 GlyPheHisThrAlaSerGlyLysLysIleGluLeuThrLysGlu 901  
QY 1519 TGTACTCTGAATCGCGCAAACTTTATAGTTTACTCAAGCATCTTAACTCTGCCATAGC 1578  
Db 902 SerLeuAspLysValLysAsnLeuPheGluGluLysGluGlnAspAsnSer 918  
QY 1579 AGATTATCTCAACTATGTTTATCATCCAAAGGTGCTAGATAGATGTTTCTATCAAT 1638  
Db 919 GluIleThrAsnPheSerHis-----ArgGlyAlaLysMetSerLysAspArgGlu 935  
QY 1639 GAGTGTATATGAGGCG---TCCTATCGAGAAATCCTCAGGATATTCATCGCAACCTGGA 1695  
Db 936 GluCysLysAspGlyArgGluLeuAlaCysGlyThrThrGluIleThrThrProGlu 955  
QY 1696 TTTGCTTTTAGTCGCAAGGACAGTAAAGAGAACCTATCACATATCTTGTGTGC 1755  
Db 956 TyrGluGluThrHisSerSerLeuGluLysLysValSerAsnGluIleAlaAlaLeu 975  
QY 1756 AGGCCA-----AAACGAACA----- 1770  
Db 976 ArgProArgLeuLeuSerAspAsnLeuTyrLysGlnThrGluAsnLeuLysIleSerAsp 995  
QY 1771 -----AAAGCAAGCATGCTCGAATTTCTTGAATCTGAAGTGGGAA----- 1812  
Db 996 HisAlaSerGlnLysValAspValHisGluAsnThrGluLysGluThrAlaLysLysPro 1015  
QY 1813 -----GTAGACACAAAGACATATAGTAGTGGCCACATCGT---CTGTATTTCCAT 1863  
Db 1016 ThrMetTyrThrAsnGlnSerThrTyrSerAlaIleGluAsnSerProLeuThrPheThr 1035  
QY 1864 AGTGATACCTGCTTACCTCTCGTCCACAGAAATGGAAGTAGATAGTGAAGATGAA--- 1920  
Db 1036 GlnAspThr-----GluGluLysPheSerValSerGluAlaSerLeu 1049  
QY 1921 ---AAGGATCTGATGCTGAAGAGAAACCAATACAAATGAAGATTTCTTGAT 1977  
Db 1050 PheGluAlaLysLysTrpLeuArgGlu-----GlyGluTrpAspAsp 1063  
QY 1978 -----GTTAATGAAGGAGAGAAAGAGATGATGAATCTCTGGAATCTCCATGTC 2025  
Db 1064 GlnSerGluArgIleAsnAlaLysValAsnCysLysGluTyrProAspAspTyrVal 1083  
QY 2026 -----ATGAAGCATGGTTTATGCTGACATCAATCAATGAATCAT 2064  
Db 1084 GluAsnProSerCysGlyAsnSerSerAsnSerAlaIleThrGluAsnAspLysAsnHis 1103  
QY 2065 GCCTGTATGCTGTTGTAGAAATATGACAGAAATAATTAAGAAGATTTATGTCGA 2124  
Db 1104 -----LeuSerGluLysGlnGlySerThrTyrLeuSerAsnSerThrMetSer 1119  
QY 2125 AAC---TTTCATGCTTCTATGCTAGTCAGCATGATGAC-----TTTAATCTTATTAGC 2172  
Db 1120 AsnSerTyrSerTyrHisProGlyPheCysHisSerSerGluValTyrAsnLysSerGlu 1139  
QY 2173 ATAATGTC-----ATAGATAAGCTGTTACCAAGCTCGT 2208  
Db 1140 TyrLeuSerArgSerLysIleAspAsnSerGlyIleGluProValIleLysAsnIleArg 1159  
QY 2209 GAAATGCA----- 2217  
Db 1160 GluArgLysAsnIleGlyPheSerGluIleMetSerProGlyArgGluAlaAspThrAsp 1179  
QY 2218 -----CAAAAATTAGAAAGCGGGAATCTGCTCCCT 2250  
Db 1180 ProGlnSerValAsnAspIleCysValGluLysLeuAlaThrAsnSerSerCysLysAsn 1199  
QY 2251 GCAAAC----- 2262  
Db 1200 LysAsnThrAlaIleLysValAlaIleSerAspSerAsnAsnPheAsnThrIleGlnLys 1219

QY 2263 ATAACCTGAAGAACAAATGGGACAGCAATGATGTTAGTGAATTAACCTCAAAAGAG--- 2319  
Db 1220 LeuAsnSerAspSerAsnAsnSerValProAlaTyrSerThrValAsnSerLysArgVal 1239  
QY 2320 -----AAAGCTTTGGAA-----ACAGATAGTGTCTCAGGGTTTCA 2355  
Db 1240 PheValAlaHisGlnThrLysValThrGluGlyPheThrAspAsnCysSerMetValThr 1259  
QY 2356 AAACAGACAGCAAAACAAACAAA 2376  
Db 1260 LysGlnAsnThrLysSerLys 1266  
RESULT 7  
US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5080  
; LENGTH: 3696  
; TYPE: PR1  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5080  
Alignment Scores:  
Pred. No.: 8,67e-06 Length: 3696  
Score: 156.50 Matches: 140  
Percent Similarity: 35.74% Conservative: 147  
Best Local Similarity: 17.43% Mismatches: 327  
Query Match: 2.04% Indels: 189  
DB: 4 Gaps: 32  
US-09-874-162A-7 (1-4409) x US-09-134-001C-5080 (1-3696)  
QY 139 ATCGACACATCGAGACACCAACATCATGATCAGATCCACGGTTTTAGAAAAACAGAA 198  
Db 1926 LeuGluLysIleAsnAsnIleGlnProSerThrGlnThrLysThrAsnAlaLysGlnGlu 1945  
QY 199 TTACAGACACCAACCTATGTTCCCTGAGTTACATAAATAGATTTCATCAGATGCTGCC 258  
Db 1946 IleAsnAspLysAlaGlnGluIleGlnIleAsnAsn---ThrProAspAlaThr 1964  
QY 259 CCGCAGACAGAGTCCTCTAAAGAGAGATTCAGCCGAGAGCTCTCGTGAATCTGCTGCC 318  
Db 1965 GluGluGluLysGlnGluAlaThrAsnArgValAsnAlaGlyLeuAlaGlnAlaIleGln 1984  
QY 319 AGCTCAGTCTCGAGGAATGTCCTCCACTCCGCCACCCACGAGTGGAGCCCTTACT 378  
Db 1985 -----AsnIleAsnAsn-----AlaHisSerThrGlnGluValAsn 1996  
QY 379 CCCCCCGTAGCCCCCACCCTCCCTCTTTCATTCGCGACGACGACCTCCGACAGAG 438  
Db 1997 GluSerLysThrAsnSerIleAlaThrIleLysSerValGlnProAsnValIleLysLys 2016  
QY 439 CCAACACAGATCTATAGATTTCTTCCGAACTCGGATCTCATAGCACCATAATTTTGCAC 498  
Db 2017 ProThrAlaIleAsnSerLeuThrGlnGluAlaAsn----- 2028  
QY 499 AGAAGCTTCTTACTTACTGTTCTCATCGAACTCCAGAACTCCAGAAACATCAAAAGGAA 552  
Db 2029 AsnGlnLysThrLeuIleGlyAsnAspGlyAsnAlaThrAspGluLysGluAlaAla 2048

QY	553	-----ACATTAAAGTTGATGATATGTTATCAAAAGTAGAGAAAAATGAAGGA	600
Db	2049	LySGlnLeuValThrGlnLysLeuAsnGluGlnIleGlnLysHisGluSerThrGln	2068
QY	601	GAGCAAGAATCTCATAGCTTCAGTCAT-----TTTGACGCTTACGTTTACT	648
Db	2069	AspAsnGlnValAspAsnValLysAlaGlnAlaIleThrAlaIleLysLeuIle	2086
QY	649	GGTTTCTTCCAAAAATGATAAGCCATCAACCAACTCAGAAAAATGAACAAAATCTGTGT	708
Db	2087	-----AsnAlaAsnAlaHisLysArgGlnAspAlaIle	2097
QY	709	ACCCTGGAAGTCTGCTGTGCAAGTTTGCCACAAAAAGAAAGAGTAAGTTGTGCCA	768
Db	2098	Asn-----IleLeuThrAsnLeuAlaGluSerLysSerAspIleArgAla---	2113
QY	769	ATAAGCGAAGTTCACACAGGTAAAGACGAGTGCCTTTGATTCCTGACCTCAATCAACA	828
Db	2114	--AsnGlnAspAlaThrGluGluAsnThrAlaIleGlnSerIleAspAspThr	2132
QY	829	AAACCCGGAATTTCCCGTCCCTTCAGTTTCCAGTAATGAATTTGAACCTAGTAACAGC	888
Db	2133	-----LeuAlaGlnAlaArgAsnAsnIleAsnGlyAlaAsnThr	2145
QY	889	CAVATGCTGAAGTCTTACTCGTCTATTAGTAGTACTCGTCAGAGCAAGACAGAGTTT	948
Db	2146	AsnAlaLeuValAspGluAsnLeu-----GluAspGlyLysGlnLysLeu	2160
QY	949	AATGGAATGATTAAATGGA-----	966
Db	2161	GlnArgIleValLeuSerThrGlnThrLysThrGlnAlaLysAlaAspIleAlaGlnAla	2180
QY	967	-----GAAACCAATGAAAATATTGATGTCATGCAATGAAGAGCTT	1002
Db	2181	IleGlyGlnGlnArgSerThrIleAspGlnAsnGlnAsnAlaThrThrGluGluLysGln	2200
QY	1003	CCAGCCAGAGAAACGAAATTCGTGAG-----GATGGGGAACACATTT	1047
Db	2201	GluAlaLeuGluArgLeuAsnGlnLysThrAsnGlyValAsnAspArgIleGlnAlaAla	2220
QY	1048	GTTCACAAATGACAGTATTGTAT-----AAAAACAGGCGCTTACAGCTTTTAGATGGGAA	1104
Db	2221	LeuAlaAsnGlnAsnValThrAspGluLysAsnAsnIleLeuGluThrIleArgAsnVal	2240
QY	1105	TATGAAGTACCCATCAGAGAAATGGAGAATGTCCATAAGACGAGAAAGACCAATGG	1164
Db	2241	GluProIleValIleValLysProLysAlaAsnGluIleIleArgLysLysAlaAlaGlu	2260
QY	1165	GAGACTATTCTTGATGGGAAGAGCTCCCTCCATTCGAAACATTTCTCAGGACCTPAG	1224
Db	2261	GlnThrThrLeuIleAsnGln-----AsnGlnAspAlaThr	2272
QY	1225	TTG-----CAGTTTCACTCTCGTTGACAGGAGACCAATGATAATCTACGGCT	1275
Db	2273	LeuGluGluLysGlnIleAlaLeuGlyLysLeuGluGluValLysAsnGluAlaLeuAsn	2292
QY	1276	CTTATGCCAAACCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAAAAC-----	1329
Db	2293	GlnValSerGln-----AlaHisSerAsnAsnAspValLysIleAlaGluAsnAsnGly	2310
QY	1330	-----AAGCTGTCTTCAGTTAAACCTACTCAAACTATTGCT-----GTTAAGAGATCA	1377
Db	2311	IleAlaLysIleSerGluValHisProGluThrIleIleLysArgAsnAlaLysGlnGlu	2330
QY	1378	TTGACTACAGATCTTACAAACAGAAAAAGGATACCTCCAATGAAAAACCGACAAAA	1437
Db	2331	IleGluGlnAspAlaGlnSer-----GlnIleAspThrIleAsnAlaAsn-----	2345
QY	1438	TTAAGATATTATCAGTTTCTCTATAACAACAATACAGGCAACAACTGAAGCAGA	1497
Db	2346	-----AsnLysSerThrAsnGluLysSerAlaAla	2356
QY	1498	GATGACCTGATTGCCCTTGGTGTACTCTGAACTCGCGCAAACTTTATAGTTTACTCAAG	1557

Db	2357	IleAsp	-----ArgValAsnValAlaLysIleAspAlaIle-----	2368
Qy	1558	CATCTTAAACTGCGCATAGACAGATTTCCTCAACTATGCTTTATCATCCAAAAGGTGCT	1617	
Db	2369	-----AsnAsnIleThrAsnAlaThrThrThr	2377	
Qy	1618	AGGATAGATGTTCTTATCAATGAGTGTTATGATGGCTCTTATGATGAGAAATCTCCAGGAT	1677	
Db	2378	GlnLeu-----ValAsn-----AspAlaLysAsnSerGlyAsnThrSerIle	2391	
Qy	1678	ATTCATCGCAACTCGATTTGCTTTAGTCGCAACGGACCACTTAAAGAAACACCTATC	1737	
Db	2392	SerGlnIleLeuProSerThrAla-----Vallys-----	2401	
Qy	1738	ACACATATTTCTGTGTGCGAGCCCAAAACGAAACCAAGCATGCTGTAATTTCTTGAA	1797	
Db	2402	ThrAsnAlaLeuAlaAlaLeuAlaSerGluAlaLysAsnLysAsnAlaIleIleAspGln	2421	
Qy	1798	TCGTGAAGATGGGGAAGTAGAACACAGCAAGAACATATAGTAGTGCCACATCTCTGTAT	1857	
Db	2422	ThrProAsnAlaThrAlaGluGluLysGluGluAlaAsnAsnLysValAspArgLeu---	2440	
Qy	1858	TTCCATAGTGTATACCTGCTTACCTCTCCGTCACAAAGAAATGGAAGTAGATAGTGAAGAT	1917	
Db	2441	-----GlnGlu	2442	
Qy	1918	GAAAGAGATCCTGAATGGCTTAAGAGAAAAACCATTTACACAAAT	1962	
Db	2443	GluAlaAspAlaAsnIleLeuLysAlaHisThrThrAspGluValAlaAsnAlaIleLysAsn	2462	
Qy	1963	GAAGAGTTTCTCATGTGTTAATGAAGGAGAAAGAGTGTGATAAATCTCGAATCTCCAT	2022	
Db	2463	GlnAlaValGlnAsnIleAsnAlaValGlnValGluValIleLysLysGlnAsnValLys	2482	
Qy	2023	GTCATGAAGCATGGGTTTATTTGCTGCAATCAAAATGAATCATCCCTGTATGCTGTTGTA	2082	
Db	2483	AsnGlnLeuAsnGlnPheIle---AspAsnGlnLysLys-----IleIle	2496	
Qy	2083	GAAATTTATCGACAGAAAAATAATTAACAGAGAAATTTATGTCGAACTTCATGCTTCATCTA	2142	
Db	2497	GluAsnThrProAspAlaThrLeuGluGluLysAlaGluAlaAsnArgLeuLeuGlnAsn	2516	
Qy	2143	GTC-----AGCATGCATGACTTTTAACTCTTATTAGCAATAATG	2178	
Db	2517	ValLeuThrSerThrSerAspGluIleAlaAsnValAspHisAsnAsnGluValAspGln	2536	
Qy	2179	TCATAGATAAAGCTGTTACCAGCTCCGTGAAATGCAGCAAAAATTAGAAAAGGGGAA	2238	
Db	2537	AlaLeuAspLysAlaArgProLysIleGluGluIleValProGlnValSerLysLysArg	2556	
Qy	2239	TCGTCTCCCTCCAAACGAAAGAAATAACTGAAACAAATAATGGGACAGCAAAATGGATTT	2298	
Db	2557	AspValLeuAsnAlaIleGlnGluAlaPheAsnSerGlnThrGlnGluIleGlnGluAsn	2576	
Qy	2299	AGTGAATTAACTCAAAAGAAAGCTTTGGAAACAGATAGTGTCTCAGGGGTTTCAAAA	2358	
Db	2577	GlnGluAlaThrAsnGluGluLysThrGluAlaLeuAsnLysIleAsnGlnLeuLeuAsn	2596	
Qy	2359	CAGAGCAA 2367		
Db	2597	GlnAlaLys 2599		
RESULT 8				
US-09-572-191-2				
; Sequence 2, Application US/09572191				
; Patent No. 6355466				
; GENERAL INFORMATION:				
; APPLICANT: Beraud, Christophe				
; APPLICANT: Sakowicz, Roman				
; APPLICANT: Wood, Kenneth				
; TITLE OF INVENTION: NO. 6355466el motor proteins and methods for				
; TITLE OF INVENTION: their use				

RESULT 8  
US-09-572-191-2  
; Sequence 2, Application US/09572191  
; Patent No. 6335466  
; GENERAL INFORMATION:  
; APPLICANT: Seraud, Christophe  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Wood, Kenneth  
; TITLE OF INVENTION: NC. 63554661  
; TITLE OF INVENTION: their use

```

; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572.191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2

Alignment Scores:
Pred. No.:      8,4e-06      Length:      1388
Score:          154.50      Matches:      184
Percent Similarity: 36.14%      Conservative: 129
Best Local Similarity: 21.25%      Mismatches: 306
Query Match:      2.01%      Indels:      248
DB:              4          Gaps:      40

US-09-874-162A-7 (1-4409) x US-09-572-191-2 (1-1388)
QY 169 ACAGATCCACCGGTTTTAGAA---AAACAAGATTACAGCAGCCAACTATGTTGCCCTG 225
Db 169 ACAGATCCACCGGTTTTAGAA---AAACAAGATTACAGCAGCCAACTATGTTGCCCTG 225
QY 609 ThrArgLysArgGlnLeuGluSerGluLeuGlnSerLeuGlnLysAlaAsnLeu 628
Db 609 ThrArgLysArgGlnLeuGluSerGluLeuGlnSerLeuGlnLysAlaAsnLeu 628
QY 226 AGTTACATAAATAGATTC---ATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAG 282
Db 226 AGTTACATAAATAGATTC---ATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAG 282
QY 629 AsnLeuGluAsnLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu 648
Db 629 AsnLeuGluAsnLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu 648
QY 283 AAGAGATTTCAGCCGAACTCTCGTGACTCTGCTCCAGCTCAGTGTCTCGAGGGAATGTG 342
Db 283 AAGAGATTTCAGCCGAACTCTCGTGACTCTGCTCCAGCTCAGTGTCTCGAGGGAATGTG 342
QY 649 AsnLysIleHisAlaGlu-----ThrLeuLysIle-----Ile 659
Db 649 AsnLysIleHisAlaGlu-----ThrLeuLysIle-----Ile 659
QY 343 TCCACTCCCCACGC-----CAGAGAGTGGAGACCTTACTCCCGCGTGACC 390
Db 343 TCCACTCCCCACGC-----CAGAGAGTGGAGACCTTACTCCCGCGTGACC 390
QY 660 ThrThrProThrLysAlaThrGlnLeuHisSerArg-----ProVal--- 673
Db 660 ThrThrProThrLysAlaThrGlnLeuHisSerArg-----ProVal--- 673
QY 391 CCACCCATCACCCCTCC---TCTTCATTCCGACGACGACTCCGACAGAGCCCAACACAG 447
Db 391 CCACCCATCACCCCTCC---TCTTCATTCCGACGACGACTCCGACAGAGCCCAACACAG 447
QY 674 ProLysLeuSerProGluMetGlySerPheGlySer----- 685
Db 674 ProLysLeuSerProGluMetGlySerPheGlySer----- 685
QY 448 ATCTATAGATTTCCTGAACTCGGAAT-----CTCATAGACCAATATTTTCACAGA 501
Db 448 ATCTATAGATTTCCTGAACTCGGAAT-----CTCATAGACCAATATTTTCACAGA 501
QY 686 -----LeuTyThrGlnAsnSerSerIleLeuAspAsnAspIleLeuAsnGlu 701
Db 686 -----LeuTyThrGlnAsnSerSerIleLeuAspAsnAspIleLeuAsnGlu 701
QY 502 ACTCTTACTTACATGCTCTCATCGAACTCCAGAACACATCAAAAGGAAACATTAAA 561
Db 502 ACTCTTACTTACATGCTCTCATCGAACTCCAGAACACATCAAAAGGAAACATTAAA 561
QY 702 ProValProProGluMetAsnGluGlnAlaPheGluAlaIleSerGluGluLeuArgThr 721
Db 702 ProValProProGluMetAsnGluGlnAlaPheGluAlaIleSerGluGluLeuArgThr 721
QY 562 GTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGAGCAAGAAATCTCATAGCTTG 621
Db 562 GTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGAGCAAGAAATCTCATAGCTTG 621
QY 722 ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGluGlu----- 737
Db 722 ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGluGlu----- 737
QY 622 TCAGTCTATTTCGACCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCACCA 681
Db 622 TCAGTCTATTTCGACCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCACCA 681
QY 738 -----HisLysAsnLeuLysLeuGlnGln 745
Db 738 -----HisLysAsnLeuLysLeuGlnGln 745
QY 682 AACTCAGAAAAT---GAACAAAATTCGTTTACCCTGGAAGTCCTGTTGCAAGATTTCG 738
Db 682 AACTCAGAAAAT---GAACAAAATTCGTTTACCCTGGAAGTCCTGTTGCAAGATTTCG 738
QY 746 HisValAspLysLeuGluHisSerThrGlnMetGlnGluLeuPheSerSer----- 763
Db 746 HisValAspLysLeuGluHisSerThrGlnMetGlnGluLeuPheSerSer----- 763
QY 739 CACAAAAAAGAAAGATGTAAAGTTGTCCTCAATAGGCAAGTCCACAGAGTAAGAACAG 798
Db 739 CACAAAAAAGAAAGATGTAAAGTTGTCCTCAATAGGCAAGTCCACAGAGTAAGAACAG 798
QY 764 -----GluArgIleAspTrpThrLysGlnGln 772
Db 764 -----GluArgIleAspTrpThrLysGlnGln 772
QY 799 GTGCTTTGATTCTCGACTCAAT-----CAACAAAA 831
Db 799 GTGCTTTGATTCTCGACTCAAT-----CAACAAAA 831
QY 773 GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnThrGlnThrLys 792
Db 773 GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnThrGlnThrLys 792
QY 832 CCCGGAATTTCCCGTCCCTTGCAGTTTCCAGTAAATGAATTTGAACCTAGTAACAGCCAT 891
Db 832 CCCGGAATTTCCCGTCCCTTGCAGTTTCCAGTAAATGAATTTGAACCTAGTAACAGCCAT 891
QY 793 -----AsnAspPheLeuLysSerGluValHis 801
Db 793 -----AsnAspPheLeuLysSerGluValHis 801

892 -----ATGCTGAAGTCTTACTCG 909
802 AspLeuArgValValLeuHisSerAlaAspLysGluLeuSerSerValLysLeuGluTyr 821
910 TTGCTATTATAGAGTACGCTCCAGGAAGAGAGAGAGTAAATGGAATGATTAAATCGAGAA 969
822 SerSerPheLysThrAsnGln-----GluLysGluPheAsnLysLeu----- 835
970 ACCAATGAAATATTGATGCTCAATGAAGAGCTTCAGCCAGAGAAACGAAATCGTGAG 1029
836 SerGluArgHisMetHisValGlnLeuGlnLeuAspAsnLeuArgLeuGluAsnGluLys 855
1030 GATGGGAAACACATTTCTGCACAAATGACAGTATTGTATAAAACAGCGCTTACAG 1089
856 LeuLeuGluSerLys--AlaCysLeuGlnAspSerTyAspAsnLeuGlnGluIleMetL 875
1090 CTTTATAGATGGGAAATATCAAGTAGCCATGAGGAAATGGAAGATGTCCCAATAGCAAG 1149
875 yspheGluIleAspGlnLeuSerArgAsnLeuGlnAsnPheLysLysGluAsnGluThrL 895
1150 AAAAGACCAATCGGAGACTATTCTTGTATGGGAAGAGCTGCCCTCATTTCGAACATTT 1209
895 euLysSerAspLeuAsnAsnLeuMetGluLeuGluAlaGluLysGluArgAsn--- 913
1210 TCTCAGGACCTACGTTGCTGAGTTCTCTCTCGTCGACGAGAGACCAATGATAAATCT 1269
914 --AsnLysLeu--SerLeuGlnPheGlu-----GluAspLysGluAsnSer 927
1270 ACGGCTCTTATTCGCAAACTCTTCCCTACTAGAAATTCAGAGAGTCTCCATCAGGAAAC 1329
928 SerLysGluIleLeuLysValLeu-----GluAlaValArgGlnGluLys 942
1330 AAGCCTGGTTCAGTTAAACCTACTCAAACTATTGCT-----GTTAAAGAAATCA 1377
943 GlnLysGluThrAlaLysCysGluGlnGlnMetAlaLysValGlnLysLeuGluGlnSer 962
1378 TTGACTTACAGATCTCAAAACAGAAAA-----GAAAGAGATCTCCAAAT 1422
963 Leu-----LeuAlaThrGluLysValIleSerSerLeuGluLysSerArgAspSer 979
1423 GAAACACCGA-----CAAAATTAAGAAATATTTATCAG 1455
980 AspLysLysValValAlaAspLeuMetAsnGlnIleGlnGluLeuArg----- 995
1456 TTTCTCTATAACAACAATAACAGGCAACAACTGAAGCAAGAGATGACTGCTGCCT 1515
996 -----SerSerValCysGluLysThrGluThrIleAspThrLeuLysGlnGlu 1011
1516 TGGTGACTCTGAATGCTCGGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCGCCAT 1575
1012 LeuLysAspIleAsnCysLys-----Tyr--AsnSerAlaLe 1023
1576 ACCAGATTATCTTCACTTATCATGTTTATCATCAAAAGTGCTAGGATAGATGTTCTATC 1635
1023 uValAspArgGluGluSerArgValLeuIleLysLys----- 1035
1636 AATGAGTGTATGATGGCTCTCTATCAGGAAATCTCAGGATATTCATCGCAACCTGGA 1695
1036 -GlnGluValAspIleLeuAspLeuLysGluThrLeuArgLeuArgIleLeuSerGluAs 1055
1696 TTTGCTTTTAGTCGCAACGGACCTTAAGAGAACCTATACACATATACACATA----- 1744
1055 p-----IleGluArgAspMetLeuCysGluAspLeuAlaHisAlaThrGluGlnLe 1072
1745 -----TTCTTGTGTCAGCGCCAAAACGA 1767
1072 uAsnMetLeuThrGluAlaSerLysLysHisSerGlyLeuLeuGlnSerAlaGlnGlu 1092
1768 A-----CAAAGACAGCATGCTCAATTTCTGAATCTGAAGATGG----- 1809
1092 uLeuThrLysLysGluAlaLeuIleGlnGluLeuGlnHisLysLysAsnGlnLysLysG 1112
1810 -GAAGTAGAACCAAGCAAGAACA---TATAGTAGTGGCCACAATCGTCTG----- 1854

```

Alignment Scores:		8.4e-06	Length:	1388
Pred. No.:	Score:	154.50	Matches:	184
	Percent Similarity:	36.14%	Conservative:	129
	Best Local Similarity:	21.25%	Mismatches:	306
	Query Match:	2.01%	Indels:	248
DB:		4	Gaps:	40

US-09-874-162A-7 (1-4409) x US-09-723-262-2 (1-1388)

QY	169	ACAGATCCACGGGTTTTAGAA---	AAACAAGAAATTACAGCAGCAGCAACCTATGTTGCCCTG	225
Db	609	ThrArgLysArgGlnLeuGluLeuGlnSerGluLeuGlnSerLeuGlnLysAlaAsnLeu	628	
QY	226	AGTTACATAAATAGATTC---	ATGCACAGATGCTGCCGCCGAGCAGCAGGAGTCCCTAAAG	282
Db	629	AsnLeuGluAsnLeuLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu	648	
QY	283	AAGNAGATTACGCGGAGCTCTCGCTGACTCTCTCCAGCTCAGTGTCTCAGGGAATGTG	342	
Db	649	AsnLysIleHisAlaGlu-----	ThrLeuLysIle-----	659
QY	343	TCCACTCCCCCAGC-----	CACAGCAGTGGAGCCTTACTCCCCCGTGACC	390
Db	660	ThrThrProThrLysAlaTyrGlnLeuHisSerArg-----	ProVal-----	673
QY	391	CGACCCATCACCCGCTCC---	TCTTCATTCCGACAGCACTCGACAGAGCAACACAG	447
Db	674	ProLysLeuSerProGluMetGlySerPheGlySer-----		685
QY	448	ATCTATAGATTTCTTCGAATCCGAAT----	CTCATAGCACCAATATATTTTGCACAGA	501
Db	686	-----	LeuTyrThrGlnAsnSerSerIleLeuAspAsnApIleLeuAsnGlu	701
QY	502	ACTCTTACTTACATGTCATCGAAACTCGACAAACAAATCAAAAGGAAACATTTAAA	561	
Db	702	ProValProProGluMetAsnGluGlnAlaPheGluAlaIleSerGluGluLeuArgThr	721	
QY	562	GTTGATGATATGTTACAAAAGTAGAAMAATGAAAGGACAGCAAGAATCTCATAGCTTG	621	
Db	722	ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGlu-----		737
QY	622	TCAGCTCATTTCAGCTTACGTTTACTGGTTTCTTCCACAAAAATGATAAGCCATCACC	681	
Db	738	-----	HisLysAsnLeuLysLeuGlnGln	745
QY	682	AACTCAGAAAT---GAACAAAATTCTGTTACCTCGAAGTCTGCTGTTGAAAGTTTC	738	
Db	746	HisValAspLysLeuGluHisHisSerThrGlnMetGlnGluPheSerSer-----		763
QY	739	CACAAAAAAGAAAGATGTAACTTGTCACAAATGAAGCAAGTCCACAGATAAAAGCAG	798	
Db	764	-----	GluArgIleAspTrpThrLysGlnGln	772
QY	799	GTGCTCTTGATTCCTGACCTCAAT---	CAACAAAA	831
Db	773	GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnThrGlnThrLys	792	
QY	832	CCCGGAAATTCCTCGCTCCCTTGCCAGTTTCAGTAATGAATTTCAACTAGTAACAGCAT	891	
Db	793	-----	AsnAspPheLeuLysSerGluValHis	801
QY	892	-----	ATGTGGAAGTCTTACTCG	909
Db	802	AspLeuArgValValLeuHisSerAlaAspLysGluLeuSerSerValLysLeuGluTyr	821	
QY	910	TTGCTATTTAGATGACTCTCTCCAGGAAGAAGAGAGTTTAATGGAATGATTAATGGAGAA	969	
Db	822	SerSerPheLysThrAsnGln-----	GluLysGluPheAsnLysLeu-----	835
QY	970	ACCAATGAAATATTTGATGTCATTAAGAGCTTTCCAGCCAGAGAAACAAATCGTGAG	1029	

[illegible]

```

Db      836 SerGluArgHisMetHisValGlnLeuGlnLeuAspAsnLeuArgLeuGluAsnGluLys 855
QY      1030 GATGGGAAAAACATTTGTTGCAAAATGACAGATATTGTTATATAAAACAGCGCGTTACAG 1089
Db      856 LeuLeuGluSerLys--AlaCysLeuGlnAspSerTyrAspAsnLeuGlnGluLeuMetL 875
QY      1090 CTTTATAGATGGGATATGACATGACCATGAGGAATGGAAGATGTCACATGAAGCAAG 1149
Db      875 ysPheGluLeuAspGlnLeuSerArgAsnLeuGlnAsnPhelysLysGluAsnGluThrL 895
QY      1150 AAAAGAGCAACATGGAGACATTTCTTGATGGAGAGGCTGCGCTCCATTCGAAACATTTT 1209
Db      895 euLysSerAspLeuAsnLeuMetGluLeuLeuGluAlaGluLeuGluArgAsn--- 913
QY      1210 TCTCAGGACCTACGTTTCAGTTCTACTCTCTGTTGGACAGAGACCAATGATAATCT 1269
Db      914 --AsnLysLeu--SerLeuGlnPheGlu-----GluAspLysGluAsnSer 927
QY      1270 ACGGCTCTATTGCGCAACCTCTTCCCACTAGAAATTCAGAGAGTCTCCATCAGAAAAC 1329
Db      928 SerLysGluLeuLeuLysValLeu-----GluAlaValArgGlnGluLys 942
QY      1330 AAGCCTGTTGTTAAACCTACTCTCAACTATTGCT-----GTTAAAGATCA 1377
Db      943 GlnLysGluThrAlaLysCysGluGlnGlnMetAlaLysValGlnLysLeuGluGluSer 962
QY      1378 TTGACTACAGATCTACAAACAGAAAA-----GAAAGAGTACTCCAAAT 1422
Db      963 Leu-----LeuAlaThrGluLysValLysSerSerLeuGluLysSerArgAspSer 979
QY      1423 GAAAAACCGA-----CAAAATTAAGAATATTTATCAG 1455
Db      980 AspLysLysValValAlaAspLeuMetAsnGlnLeuGlnLeuArg----- 995
QY      1456 TTTCTCTATACAACTACAGGCAACAACTGACGAGATGACCTGCTGTCATGCGCCT 1515
Db      996 -----SerSerValCysGluLysThrGluThrLeuAspThrLeuLysGlnGlu 1011
QY      1516 TGGTGTACTCTGAAGTGGCGCAAACTTTATGTTTACTACAGCATCTTAAACTCTGCCAT 1575
Db      1012 LeuLysAspLysAsnCysLys-----Tyr--AsnSerAlaLe 1023
QY      1576 AGCAGATTATCTCAACTATGTTTATCATCAAAAGGTGCTAGATAGATGTTTCTATC 1635
Db      1023 uValAspArgGluGluSerArgValLeuLeuLysLys----- 1035
QY      1636 AATGAGTGTATGATGGCTCTATGAGGAATCTCAGGAATCTCAGGATATTCGCGCACTGGA 1695
Db      1036 -GlnGluValAspLysLeuAspLysGluThrLeuArgLeuArgLeuSerGluAs 1055
QY      1696 TTTGCTTTTATGCGCAACGACCGACCTTAAGAGACACTATCACATA----- 1744
Db      1055 p-----IleGluArgAspMetLeuCysGluAspLeuAlaHisAlaThrGluGlnLe 1072
QY      1745 -----TTCTGTGTGCGCGCCAAACGA 1767
Db      1072 uAsnMetLeuThrGluAlaSerLysLysHisSerGlyLeuLeuGlnSerAlaGlnGluL 1092
QY      1768 A-----CAAAAGCAAGCATCTCTGAATTTCTTGAATCTGAAGATGGG----- 1809
Db      1092 uLeuThrLysLysGluAlaLeuLeuGlnGlnHisLysLeuAsnGlnLysLysG 1112
QY      1810 -GAAGTAGACAGCAAAAGCA-----TATAGTAGTGGCCACCAATCGTCTG----- 1854
Db      1112 uGluValGlnGlnLysLysAsnGluTyrAsnPheLysMetArgGlnLeuGluHisValMe 1132
QY      1855 -----TATTTCCATGTGATAC 1871
Db      1132 tAspSerAlaAlaGluAspProGlnSerProLysThrProHisPheGlnThrHisLe 1152
QY      1872 CTGCTTACTCTCCGTCACAGAAATGGAAGTAGATGATGAAATGAAAGATCTCTGA 1931
Db      1152 uAlaLysLeuLeuGluThrGlnGlnGlnLeu-----GluAspGlyArgAlaSerLy 1170

```

```

QY      1932 ATGGCTAAGAGAAAAACCATTTACAAATTCAGAGATTTTCTGATGTTAATGAAGGAGA 1991
Db      1170 sThrSerLeuGluHisLeuValThrLysLeuAsnGluAspArgGluValLysAsnAlaG 1190
QY      1992 G-----AAAGAAGTGCATGAATCTCGAATCTCCATGTCATGATCATGAACATGG 2036
Db      1190 uLeuLeuArgMetLysGluGlnLeuArgGluMetGluAsnLeuArgLeuGluSerGlnG 1210
QY      2037 GTTTATGCTGCAAT----- 2052
Db      1210 nLeuLeuGluLysAsnTyrLeuLeuGlnGlnLeuAspAspLysLysArgGlnLysG 1230
QY      2053 -----CAATGAATCATGCTGATGCTGTTGTAGAAAATTTAGACACAGAAATAAT 2105
Db      1230 uAsnSerAspGlnAsnHisProAspAsnGlnGlnLeuLysAsnGlnGluGluSerIl 1250
QY      2106 TAAGAAGATTTATGTCGAACTTCATCTTCAT-----CTAGTCAGCATGCGATGACTT 2159
Db      1250 elysGluArgLeuAlaLysSerLysLeuValGluGluMetLysMetLysAlaAspLe 1270
QY      2160 TAACTTTATGACATAATGCTCAATAGATAAAGCTGTTACCAAGCTCCCT----- 2208
Db      1270 uGluGluValGlnSerAlaLeuTyrAsnLysGluMetGluCysLeuArgMetThrAspG 1290
QY      2209 -----GAAATGCAGCAAAAATTAGAA-----AAGGGGAATCTGCTTCCCC 2249
Db      1290 uValGluArgThrGlnThrLeuGluSerLysAlaPheGlnGluLysGlnLeuArgSe 1310
QY      2250 TGCAAACGAGAAAAATACGAGAACAAAAATCGGACACAAATGGAATTTAGTGAATTA 2309
Db      1310 rLysLeuGluGluMetTyrGluGluArgThrSerGlnGluMetGluMetLeuArg 1330
QY      2310 CTCAAAAGAGAAAGCTTTGGAAACAGATAGTCTCAGGGGTTTCAAAACAGACGAAAA 2369
Db      1330 gLysGlnValGluCysLeuAlaGluGluAsnGlyLysLeuValGlyHisGlnAsnLeuH 1350
QY      2370 ACAAAACTC 2379
Db      1350 sGlnLysIle 1353

```

## RESULT 10

```

US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; PRIOR FILING DATE: 2000-11-27
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-219-2

```

```

Alignment Scores:
Pred. No.: 8,4e-06 Length: 1388
Score: 154.50 Matches: 184
Percent Similarity: 36.14% Conservative: 129
Best Local Similarity: 21.25% Mismatches: 306
Query Match: 2,01% Indels: 248
DB: 4 Gaps: 40

```

US-09-874-162A-7 (1-4409) x US-09-723-219-2 (1-1388)

QY	169	ACAGATCCACGGTTTTAGAA---AAACAAGAAATTACAGCAGCCAAACCTATGTTGCCCTG	223
Db	609	ThrArgLysArgGlnLeuGlnLeuGlnSerGluLeuGlnSerLeuGlnLysAlaAsnLeu	628
QY	226	AGTTACATAATAGATTTC---ATGACAGATGTCGCCCGCAGACAGCAGAGTCCCTAAAG	282
Db	629	AsnLeuGluAsnLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu	648
QY	283	AAGAGATTACGCCGAAGCTCTCGCTGACTCTGTCTCAGCTCAGTCTCTCGAGGGAAATGTG	342
Db	649	AsnLysIleHisAlaGlu-----ThrLeuLysIle-----:::---ile	659
QY	343	TCCACTCCCCCAGCG-----CACAGCAGTGGAAAGCCTTACTCCGCCCGTACC	390
Db	660	ThrThrProThrLysAlaThrGlnLeuHisSerArg-----ProVal---	673
QY	391	CCACCCATCACCCCTCC---TC TTCATTCCGCAGCAGCAGCTCCGACAGAGCAACACAG	447
Db	674	ProLysLeuSerProGluMetGlySerPheGlySer-----	685
QY	448	ATCTATAGATTCTTCGAACTCGGAAT-----CTCATAGCACCAATATTTTTCACAGA	501
Db	686	-----LeuTyrThrGlnAsnSerSerIleLeuAspAsnAspIleLeuAsnGlu	701
QY	502	ACTCTTACTCATGTCATCGCAAACTCCAGAACAAACATCAAAAGGAACAACTTTAAA	561
Db	702	ProValProGluMetAsnGlnGlnAlaPheGluAlaIleSerGluGluLeuArgThr	721
QY	562	GTTGATGATATGTTATCAAAAGTAGAGAAAATGAAAGGAGAGCAGAGATCTCATAGCTTG	621
Db	722	ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGluGlu-----	737
QY	622	TCAGTCTATTGCGAGCTACGTTTACTGGTTCTTCTCCACAAAAATGATAGCCATCACC	681
Db	738	-----HisLysAsnLeuLysLeuGlnGln	745
QY	682	AACTCAGAAAAAT---GAACAAAAATCTCTTACCCTGGAAAGTCCTGTTGTAAGATTGC	738
Db	746	HisValAspLysLeuGluHisSerThrGlnMetGlnGluLeuPheSerSer-----	763
QY	739	CACAAAAAAGAAAGATGTAAAGTTGTCCAATAAGGCAAGTTCACACAGGTAAAAAGCAG	798
Db	764	-----GluArgIleAspTyrThrLysGlnGln	772
QY	799	GTGCTTTGATCTCTGACCTCAAT-----CAACAAAA	831
Db	773	GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnGluThrGlnThrLys	792
QY	832	CCCGGAATTTCCCGCTTGGCAGTTTCCAGTAAATGAATTTGAACCTAGTAACAGCCAT	891
Db	793	-----AsnAspPheLeuLysSerGluValHis	801
QY	892	-----ATGTGGAAGCTCTTACTCG	909
Db	802	AspLeuArgValValLeuHisSerAlaAspLysGluLeuSerSerValLysLeuGluTyr	821
QY	910	TTGCTATTAGAGTGACTCGTCCAGGAAGAGAGAGTTTAAATGCAATGATTAAATGGAGAA	969
Db	822	SerSerPheLysThrAsnGln-----GluLysGluPheAsnLysLeu-----	835
QY	970	ACCAATGAAATATTGATGTCATGACAGAGCTTCGAGCCAGAGAAAACGAAATCGTGAG	1029
Db	836	SerGluArgHisMetHisValGlnLeuGlnLeuAspAsnLeuArgLeuGluAsnGluLys	855
QY	1030	GATGGGAAAAAGACATTGTTGGCAAAATGACATTTTGTATATAAAACAGCGCGCTTACAG	1089
Db	856	LeuLeuGluSerLys--AlaCysLeuGlnAspSerTyrAspAsnLeuGlnGluIleMetL	875
QY	1090	CTTTTAGATGGGAATATGAATGATAGCCATGCAGGAATGGAAGAAATGTCCAATGAAGCAAG	1149
Db	875	ysPheGluIleAspGlnLeuSerArgAsnLeuGlnAsnPheLysGluAsnGluThrL	895

1150	QY	AAAAGACCAACATGGGAGACATCTTCTGATGGGAAGAGGCTGCCTCATTCGAAACATTT	1209
		:     : :	
895	Db	eUlySerAspLeuAsnLeuMetGluLeuGluAlaGluLysGluArgAsn---	913
		:     : :	
1210	QY	TCTCAGGACCTACGTTGCAGTTCTCACCTCTTCGTTGGACAGAGAGACCAATGATAAATCT	1269
		:   :     : :	
914	Db	--AsnLysLeu--SerLeuGlnPheGlu-----GluAspLysGluAsnSer	927
		:   :     : :	
1270	QY	ACGGCTCTATTGTCACAACTCTTCCCACTAGAAATTCAGAGAGTCTCCCATCAGGAAAC	1329
		:   :     : :	
928	Db	SerLysGluIleLeuLysValLeu-----GluAlaValArgGlnGluLys	942
		:     : :	
1330	QY	AAGCTGTGTTCCAGTTAACTTACTCAAACTATTGCT-----GTTAAAGATCA	1377
		:   :     : :	
943	Db	GlnLysGluThrAlaLysCysGluGlnGlnMetAlaLysValGlnLysLeuGluGluSer	962
		:     : :	
1378	QY	TTGACTACAGATCTACAAACAAGAAAA-----GAAAGAGATACTCCAAAT	1422
		:     : :	
963	Db	Leu-----LeuAlaThrGluLysValIleSerSerLeuGlnLysSerArgAspSer	979
		:     : :	
1423	QY	GAAAAACCGA-----CAAAANTTAAGAATATTTATCAG	1455
		:   :     : :	
980	Db	AspLysLysValValAlaAspLeuMetAsnGlnIleGlnGluLeuArg-----	995
		:     : :	
1456	QY	TTTCTCTATACAAACATACAAAGCAACAACTGAAGCAAGAGATGACCTGCATTCGCCCT	1515
		:   :     : :	
996	Db	-----SerSerValCysGluLysThrGluThrIleAspThrLeuLysGlnGlu	1011
		:     : :	
1516	QY	TGCTGTACTCTGAATCTCCGCAAACTTTATAGTTTACTCAAGCATCTTAACATCTGCCAT	1575
		:   :     : :	
1012	Db	LeuLysAspIleAsnCysLys-----TyrAsnSerAlaLe	1023
		:     : :	
1576	QY	ACGAGATTATCTTCAACTATGTTATCATCCAAAGGTCTAGGATAGATGTTCTATC	1635
		:   :     : :	
1023	Db	uValAspArgGluGluSerArgValLeuIleLysLys-----	1035
		:     : :	
1636	QY	AATGAGTGTATGATGCTCCTATCGAGAAATCTCTCAGGATATTCATGCCAACCTGGA	1695
		:   :     : :	
1036	Db	-GlnGluValAspIleLeuAspLeuLysGluThrLeuArgLeuArgIleLeuSerGluAs	1055
		:     : :	
1696	QY	TTTGCTTTTAGTCGCAACGCGACCAAGTAAAGAACACCTATCACATA-----	1744
		:   :     : :	
1055	Db	-----IleGluArgAspMetLeuCysGluAspLeuAlaHisAlaThrGluGlnLe	1072
		:     : :	
1745	QY	-----TTCTGTGTGTCAGCGCCAAACGA	1767
		:   :     : :	
1072	Db	uAsnMetLeuThrGluAlaSerLysLysHisSerGlyLeuLeuGlnSerAlaGlnGlu	1092
		:     : :	
1768	QY	A-----CAAAAGCAACATGCTCGAATTTCTTGAATCTCAAGATGGG-----	1809
		:   :     : :	
1092	Db	uLeuThrLysLysGluAlaLeuIleGlnGluLeuGlnHisLysLeuAsnGlnLysLysG	1112
		:     : :	
1810	QY	-GAAGTAGAAGACAGCAAAAGACA--TATAGTAGTGGCCACAATCTGCTG-----	1854
		:   :     : :	
1112	Db	uGluValGluGlnLysLysAsnGluTyrAsnPheLysMetArgGlnLeuGluHisValMe	1132
		:     : :	
1855	QY	-----TATTCATAGTGAATC	1871
		:   :     : :	
1132	Db	tAspSerAlaalaGluAspProGlnSerProLysThrProHisPheGlnThrHisLe	1152
		:     : :	
1872	QY	CTGCTTACCTCTCGTCCCAAGAAATGGAAGTAGATAGTGAAGATCAAAAGCATCTCGA	1931
		:   :     : :	
1152	Db	uAlaLysLeuLeuGluThrGlnGlnGluIle-----GluAspGlyArgAlaSerLy	1170
		:     : :	
1932	QY	ATGGCTAAGAGAAAAACCATTAACAAATTAAGAGAGTTTCTTGATGTTTAATGAAGGAGA	1991
		:   :     : :	
1170	Db	sThrSerLeuGluHisLeuValThrLysLeuAsnGluAspArgGluValLysAsnAlaG	1190
		:     : :	
1992	QY	G-----AAAGAAGTATGAACCTCTGMAATCTCCATGTCATGAGCATGG	2036
		:   :     : :	
1190	Db	uIleLeuArgMetLysGlnGlnLeuArgGluMetGluAsnLeuArgGluSerGlnG	1210
		:     : :	
2037	QY	GTTTATTGCTGACAAT-----	2052



Db 1210 nLeuIleGluLysThrLeuLeuGlnGlyGlnLeuAspHisLeuArgGlnLysG1 1230  
QY 2053 -----CAATGATCACTGCTGTATGCTGTTGTAGAAATATATGACAAAAATAAT 2105  
Db 1230 uAsnSerAspGlnAsnHisProAspAsnGlnGlnLeuLysAsnGlnGlnGluSerI1 1250  
QY 2106 TAAGAGAGATTATGTCGAAACTCATCTCTCAT-----CTACTGACGATGCACTT 2159  
Db 1250 eLysGluArgLeuAlaLysSerLysIleValGluMetLeuLysMetLysAlaAspLe 1270  
QY 2160 TAATCTTATTAGCAATAATGTCAATAGATAAAGCTGTTACCAAGCTCCGT----- 2208  
Db 1270 uGluGluValGlnSerAlaLeuTyrAsnLysGluMetGluCysLeuArgMetThrAspG1 1290  
QY 2209 ----GAATGACGAGAAAATTAGAA-----AAGGGGAATCTGCTTCCCT 2249  
Db 1290 uValGluArgThrGlnThrLeuGluSerLysAlaPheGlnGlnLysGluGlnLeuArgSe 1310  
QY 2250 TGCAACACGAGAAATACTGAAGAACAAATGGACACAGCAAAATGCAATTTAGTGAATAA 2309  
Db 1310 rLysLeuGluGluMetTyrGluGluArgGluArgThrSerGlnGluMetGluMetLeuAr 1330  
QY 2310 CTCAAAAGAGAAAGTTTGGAAAACAGATAGTGTCTCAGGGGTTTCAAAAACAGCAAAAA 2369  
Db 1330 gLysGlnValGluCysLeuAlaGluGluAsnGlyLysLeuValGlyHisGlnAsnLeuHi 1350  
QY 2370 ACARAACTC 2379  
Db 1350 sGlnLysIle 1353

## RESULT 11

US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 571,0022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,254  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/141,239  
; FILING DATE: 22-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-CJ 1191  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2482 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-328-254-6

Alignment Scores: 1.73e-05 Length: 2482  
Pred. No.: 152.50 Matches: 182  
Score: 34.55% Conservative: 111  
Percent Similarity: 21.46% Mismatches: 275  
Best Local Similarity: 1.98% Indels: 280  
Query Match: 1 Gaps: 44  
DB: 1  
US-09-874-162a-7 (1-4409) x US-08-328-254-6 (1-2482)  
QY 112 CTCCTACTTCCACCTGCGCAGCTCATC-----GAGCAC-----ATC 150  
Db 770 MetHisPheAlaGluLeuGlnLysPheLeuSerLeuGlnSerGluHisLysIleLeu 789  
QY 151 GAGGACACCATCATGATCCACGGGTTTAGAAAAACAA-----GAA 198  
Db 790 HisAspGlnHisCysGlnMetSerLysMetSerGluLeuGlnThrTyrValAspSer 809  
QY 199 TTACAGCAGCCCACTATGTTGCCCTGAGTTACATAAATAGATTTCATGACAGATGCTGCC 258  
Db 810 LeuLysAlaGluAsnLeuValLeuSerThrAsnLeuArgAsnPheGlnGlyAspLeuVal 829  
QY 259 CCGCGAGCAGAGAGTCCCTAAAGAAAGATTACGCCGAAGCTCTCGCTGACTGTGCC 318  
Db 830 LysGluMetGlnLeuGlyLeuGluGluValProSerLeuSer-----Ser 846  
QY 319 AGCTCAGTGTCTCGAGGGAATGTGTCCACTCCCGCCACGACGAGTGGAGGCTTACT 378  
Db 847 SerCysVal-----ProAspSerSerSerSerLeuSerLeu--- 858  
QY 379 CCCCCGTCGACCCCATCACCCTCTCTCTTCATTCCTCGCAGCAGCAGCTCCGACAG 438  
Db 859 -----GlyAspSerSerPheTyrArg-AlaLeuLeuGluGlnTh 871  
QY 439 CCAACACAGATCT---ATAGATTTCTCGAATCTCGAATCTCATAGCACCAA----- 487  
Db 871 rGlyAspMetSerLeuSerAsnLeuGluGlyAlaValSerAlaAsnGlnCysSerVa 891  
QY 488 -----TATTTTGCACAGAACTCTTACTTACATGTCTCATC-----GAAACTCCAGA 534  
Db 891 lAspGluValPheCysSerSerLeuGlnGluGluAsnLeuThrArgLysGluThrProSe 911  
QY 535 ACACATCAAAAGGAAACATTTAAAGTTGATGATATGT-----TA 576  
Db 911 rAlaProAlaLysGlyValGluGluLeuSerLeuCysGluValTyrArgGlnSerLe 931  
QY 577 TCAAAAGTAGAGAAATGAAAGGAGAGCAAGAAATCTCATAGCTTGTCTGCTCATTTGCAG 636  
Db 931 uGluLysLeuGluGluLysMetGluSerGlnGlyLeuMetLysAsnLysGluIleGlnG1 951  
QY 637 CTTACGTTTACTGGTTTCTTCCACAAAATG-----AT 669  
Db 951 uLeuGluGlnLeuSerSerGluArgGlnGlnLeuAspCysLeuArgLysGlnTyrLe 971  
QY 670 AAGCCATCACCA-----ACTCAGAAAATGAACAAATTCGTTACCTGGAGCTCTG 723  
Db 971 uSerGluAsnGluGlnTrpGlnGlnLysLeuThrSerValThrLeuGluMetGluSer-- 990  
QY 724 CTTGTGAAAGTTTGCACAAAAAGAAAGGATGTAAGTTGTCCAAATAGGCAAGTCTCC 783  
Db 991 -----LysLeuAlaAlaGluLys----- 996  
QY 784 ACAGGTAAAAGCAGGTGCTTTGATTTCTGTGACCTCAATCAACAAAAACCGGAATTTTC 843  
Db 997 -----LysGlnThrGluGlnLeuSe 1003  
QY 844 CCGTCCCTTGCAGTTTCCAGTAATGAATTTGAACCTAGTAACAGCCATATGTTGAGTCT 903  
Db 1003 r-----LeuGluLeuGluValAlaArgLeuGlnLeu--GlnGlyLeuAspLeuSer 1020  
QY 904 TACTCGTTCTATTAGATGAGTGTCTCGTCCAGGAGAGAGAGTTTTAATGGAATGATTAAT 963



Db	1021	ArgSerLeuLeuGlyIleAspThrGluAspAla	-----IleGln	1033
QY	964	GGAGAAACCAATGAAATATTGATGTCATGAAGAGCTTC	CAGCCAGAGAAACGAAAT	1023
Db	1034	GlyArg---AsnGluSerCysaspIleSerLysGlu	-----HisThr	1046
QY	1024	CGTGAAGTGGGAAAGAACACATTTGTCACAATAGCAGTATTTGATATAAAACAGCGGC	1083	
Db	1047	SerGluThrThrGluArgThr	-----ProLysHisAspVal	1058
QY	1084	TTACAGCTTTTAGTGGGAATATGAAGTAGCCATGCGAAATGGAAGATGTCCAATA	1143	
Db	1059	HisGlnIleCysAspLysAsp	-----AlaGlnGlnAspLeuAsn	1071
QY	1144	AGCAAGAAAGAGCAACATGGGACACTATCTTGATGGAGACAGCTGCCCTCCATTCGAA	1203	
Db	1072	-----LeuAsp	-----IleGlu	1075
QY	1204	ACATTTTCTCAGGACCTACCTGCAGTTCACCTCTCTGTTGGACAGAGACCACT---	1260	
Db	1076	LysIleThrGluThrGlyAlaLeuLysProThrGlyGluCysSerGlyGluInSerPro	1095	
QY	1261	GATAAATCTACGGCTCTTATTGCCAACCTCTGCCACTGAAATTCAGAGAGTCTCCAT	1320	
Db	1096	AspThrAsnTyrgluProProGlyGluAspLysThrGlnGlySerSerGluCysIleSer	1115	
QY	1321	CAG-----GAAACACAGCTCGTTTCAGTTAAACCTACTCAA	1374	
Db	1116	GluLeuSerPheSerGlyProAsnAlaLeuValPrometAspPheLeuGlyAsnGlnGlu	1135	
QY	1375	TCATTGACTACAGACTCTACAAACAGAAAGGATATCTCAAATGAAACCCGACAA	1434	
Db	1136	AspIle---HisAsnLeuGlnLeuArg--ValLysGluThrSerAsnGluAsnLeuArg	1153	
QY	1435	AAATTAAGAATATTTATCATGTTCTCTATACAACTATCAAGGCAACAACTGAAGCA	1494	
Db	1154	LeuLeuHisValIle	-----Gluasp	1160
QY	1495	AGATGACCTGCATTCGCCCTTGCTGACTCTGAACTCGCGCAAACTTTATGTTTACTC	1554	
Db	1161	Argasp-----ArgLysValGluSerLeuLeu	1169	
QY	1555	AAGCATCTAACTCTGCCATAGCAGATTTATCTTCAACTATGTTTATCATCAAAAGGT	1614	
Db	1170	AsnGluMetLysGluLeuAspSerLysLeu	-----HisLeuGlnGlu	1183
QY	1615	GCTAGGATAGATGTTCTATCAATGAGTGTATCATGGCTCCTATGTCAGGAATCTCTCAG	1674	
Db	1184	ValGlnLeuMetThrLysIleGluAlaCysIleGlu	-----Lys	1195
QY	1675	GATATTATCGCCAACTGGATTGCTTTTAGTCGAACGACCGATTAGAGAGACACCT	1734	
Db	1195	-----	-----Lys	1195
QY	1735	ATCACATATCTTGTGTGCAGCCCAAAACGACAAACAGCATGCTGTAATTTCTT	1794	
Db	1196	---LeuGluLysIleValGlyLeuLysGluAsnSerAspLeuSerGluLysLeu	1214	
QY	1795	GAACTCAAGATGGGAAGTAGAACAGCAAAACATATAGTAGTGCCCAACATCGTCTG	1854	
Db	1215	Glu-----TyrPheSerCysaspHisGlnGluLeu	1224	
QY	1855	TATTTCCATGATGATCTGTTACTCTCGTCCACAA---GAAATGGGAAGTAGATAGT	1911	
Db	1225	LeuGlnArgValGluThrSerGluGlyLeuAsnSerAspLeuGluMetHisAlaAspLys	1244	
QY	1912	GAAGATCAAAGGAT-----	-CCTGAATGGCTAAAGAA	1944
Db	1245	SerSerArgGluAspIleGlyAspAsnValAlaLysValAsnAspSerTrp---	LysGlu	1263
QY	1945	AAAACCATTCACAAATGAAGAGTTTCTGATGTTAAT	-----	1983
Db	1264	ArgPheLeuAspValGluAsnGluLeuSerArgIleArgSerGluLysAlaSerIleGlu	1283	

QY	1984	----	GAGGAGAGAGAAGATGATGAACACTCTCGAATCTCCATGTC	2025
Db	1284	HISGLUAlaLeuTyLeuGluAlaAspLeuValGlnThrGluLysLeuCysLeu	1303	
QY	2026	ATGAAGCATGGGTTTATTGCTGCACAATAAAATCAATCATGCCGTCTATGCTGCTGTTTGTAAGAA	2085	
Db	1304	GLULYS-----ASPASNGLUASNLYS-----	1310	
QY	2086	AATTATGGACAGAAAAATAATTAAAGAGAATTTATGTGAAAACTTTCATCTGCTTCATCTAGTC	2145	
Db	1311	-----GlnLysValile-----ValCysLeuGluGluGluLeuSerVaIval	1324	
QY	2146	AGCATGTCATGACTTTAAFTCTT-----ATTAGCATAATGTCCA-----	2181	
Db	1325	ThrSerGluArgASNGLINleuArgGlyLeuLeuAspThrMetSerLysLysThrThraLa	1344	
QY	2182	ATAGATAAAGTGTTTACCAAGCTCCGTGAATATGCAGCAAAAATTAGAAAAAGGGGAATCT	2241	
Db	1345	LeuAspGlnLeuSerGluLysMetLysGluLysThrGlnGluLeuGluSerHisGlnSer	1364	
QY	2242	GCTTCCCCTGCAACGAGGAATAACTGAGAACAAATGGGACAGCAAATGGATTTAGT	2301	
Db	1365	GlucysLeuHiscysIleGlnValalaglu-----Ala	1375	
QY	2302	GAATTAATACTCAAAGAGAAAGCTTTGGAAACAGATAGTGCTCAGGGGTTTCAAAA--	2358	
Db	1376	GluVallysGluLysThrGluLeuLeuGlnThrLeuSer--SerAspValSerGluLeu	1394	
QY	2359	-----CAGACCAAAAAACAAAACTCTGAATA-----	2385	
Db	1395	LeuLysAspLysThrHisLeuGlnGluLysLeuGlnSerLeuGluLysAspSerGlnAla	1414	
QY	2386	---GCTCTAACCCCATGT 2400		
Db	1415	LeuSerLeuThrLysCys 1420		
<b>RESULT 12</b>				
US-09-803-671B-2				
; Sequence 2, Application US/09803671B				
; Patent No. 6582946				
; GENERAL INFORMATION:				
; APPLICANT: WEBSTER, Marion et al				
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC				
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES				
; FILE REFERENCE: CLO01161				
; CURRENT APPLICATION NUMBER: US/09/803,671B				
; CURRENT FILING DATE: 2001-03-12				
; NUMBER OF SEQ ID NOS: 7				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 2				
; LENGTH: 1167				
; TYPE: PKT				
; ORGANISM: Human				
US-09-803-671B-2				
<b>Alignment Scores:</b>				
Pred. No.:	1,87e-05	Length:	1167	
Score:	150.50	Matches:	182	
Percent Similarity:	33.07%	Conservative:	149	
Best Local Similarity:	18.18%	Mismatches:	350	
Query Match:	1.96%	Indels:	320	
DB:	4	Gaps:	44	
US-09-874-162A-7 (1-4409) x US-09-803-671B-2 (1-1167)				
QY	273	GTCCCTAAAGNAGAGATTGACGGGAGCTCTCGTGCAC-----	311	
Db	1	MetProLysProGluArgHisAlaGluSerLeuLeuAspPileCysHisAspThrAsnSer	20	
QY	312	-----TCTGTCACGCTCACTGTCTCG	332	

Db 21 SerProThrAspLeuMetThrValThrIlysAsnGlnAsnIleLeuGlnSerIleSer 40  
QY 333 AGG-----GAATGTGCCACTCC----- 350  
Db 41 ArgSerGluGluPheAspGlnAspGlyAspCysSerHisSerThrLeuValAsnGluGlu 60  
QY 351 -----CCACG-----CCACAGCAGTGGAGCTTACTCC----- 380  
Db 61 GluAspProSerGlyGlyArgGlnAspTrpGlnProArgThrGluGlyValGluIleThr 80  
QY 381 -----CCCCGTGACCCACCCATCACCCCTCTCTTCATTCGGCAGCAGCACTCC 431  
Db 81 ValThrPheProArgAspValSerPro-----ProGlnGluMetSer 94  
QY 432 GACAGAGCCACACAGATCTATAGATTCTTCGAACTCGGATCTCATAGCACCACATATT 491  
Db 95 Gln-GluAsp-----LeuLysGluLysAsnLeuIleAsnSerSerLe 108  
QY 492 TTGTCACAGAACTCTTACTTAC---ATGCTCTCATCGAAAC-----TCCAGAACAAACAT 542  
Db 108 uGlnGluTrpAlaGlnAlaHisAlaValSerHisProAsnGluIleGluThrValGluLe 128  
QY 543 CAAAGAGAAACATTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAAGGAGA 602  
Db 128 uArgLysLysLysLeuThrMetArgProLeuValLeuGlnLysGluGluSerSerArgG1 148  
QY 603 GCAGAACTCATAGCTTGTC---GCTCATTTGCGAGCTTACGTTTAC 647  
Db 148 uLeuCysAsnValAsnLeuGlyPheLeuLeuProArgSerCysLeuGluLeuAsnIleSe 168  
QY 648 TGGTTTCTTCCACAAAATGATAGCCATCCAAACTCAGAAAATGACAAAATTCGT 707  
Db 168 rLysSerValThrArgGluAspAlaProHisPheLeuLysGluGlnArgLysSerG1 188  
QY 708 TACCTCGAAGTCCTGTTGAAAGTTGCGCAAAAAGAAAGAGATGAAGTTGTC 767  
Db 188 uGluPheSerThrSerHisMetLysTyrSerGlyArgSerIleLysPheLeuLeuProPr 208  
QY 768 AATAAGGCAAGTCCACAC-----GGTAAAGCAGGTGCTTGTGATCTCTGACCTCAA 821  
Db 208 oLeuSerLeuLeuProThrArgSerGlyValLeuThrIleProGlnAsnHisLysPhePr 228  
QY 822 TCACACAAAACCCGAAATTTCCCGTCC-----CTGCAGTTTC 860  
Db 228 oLysGluLysGluArgAsnIleProSerLeuThrSerPheValProLysLeuSerValSe 248  
QY 861 C-----AGTAATGAATTGAACCTAGTAAC----- 885  
Db 248 rValArgGlnSerAspGluLeuSerProSerAsnGluProProGlyAlaLeuValLysSe 268  
QY 885 ----- 885  
Db 268 rLeuMetAspProThrLeuArgSerSerAspGlyPheIleTrpSerArgAsnMetCysSe 288  
QY 885 ----- 885  
Db 288 rPheProLysThrAsnHisHisArgGlnCysLeuGluLysGluAsnTrpLysSerLy 308  
QY 886 -----AGCATATGTTGAAGTCTTACTCGTTGCT 914  
Db 308 sGluIleGluGluCysAsnLysIleGluIleThrHisPheGluLysGlyGlnSerLeuVa 328  
QY 915 A-----TTTAGTGTACTCTCCAGG-----AGAAGAGATTATGAAT 956  
Db 328 lSerPheGluAsnLeuLysGluGlyAsnIleProAlaValArgGluGluAspIleAspCy 346  
QY 957 GATTAAATGAGAAACCAATGAAATATTGATGTCATGAAGAGCTTCAGCAGAGAAA 1016  
Db 348 sHisGlySerLysThrArgLysProGluGluAsnSerGlnIlyrLeuSerSerArgLy 368  
QY 1017 A-----CGAAATCGTAGAGTGGGAAAGACATTGTTGACACAAAT 1058  
Db 368 sAsnGluSerSerValAlaLysAsnTyrGluGlnAspProGluIleValCysThrIlePr 388

QY 1059 GACAGATTGTTGATAAAACACAGCGCTTACAGCTTTTATGATGGGAATATGAA----- 1110  
Db 388 oSerLysPheGlnGluThrGlnHisSerGluIleThrProSerGlnAspGluGluMetAr 408  
QY 1111 -----GTAGCCATGCCAGAA----- 1125  
Db 408 gAsnAsnLysAlaAlaSerLysArgValSerLeuHisLysAsnGluAlaMetGluProAs 428  
QY 1126 -----ATGGAAGAAATGTCCAATTAAGCAAGAAACAGACCAATGGGAGACTATTCCTGA 1178  
Db 428 nAsnIleLeuGluGluCysThrValLeuLysSerLeuSerSer-----ValValPheAs 446  
QY 1179 TCGGAAGAGCGCTCCATTCGAAACATTTCTCAGGACCTACGTTGCGAGTTCACTCT 1238  
Db 446 p-----AspProIleAspLysLeuProGluGlyCysSer----- 457  
QY 1239 TCGTTGGACAGGAGACCAATGATAAA---TCTACGGCTCCTATTGCAAAACCTCTTGC 1295  
Db 458 -----SerMetGluThrAsnIleLysIleSerIleAlaGluArgAlaLysProGluMe 475  
QY 1296 CACTAGMAATTCAGAGAGTCTCCATCAGGAAACAAAGCGCTGTTTCAGTTAAACCTACTCA 1355  
Db 475 tSerArgMetValProLeuIleHis---IleThrPheProValaspGlySerProLysG1 494  
QY 1356 AACTATTGCTGTTAAAGAAATCATTTGACTACAGATCTACAAAACAGAAA----- 1404  
Db 494 uProValIleAlaLysProSer-----LeuGlnThrArgLysGlyThrIleHi 510  
QY 1405 -----GAAAGGATATCCAAATCA 1424  
Db 510 sAsnAsnHisSerValAsnIleProValHisGlnGluAsnAspLysHisLysMetAsnSe 530  
QY 1425 AAACCGA-----CAAAATTAAGAAATATTTATCATAGTTCTCTCTATAACAA 1469  
Db 530 rHisArgSerArgArgIleThrAsnLysCysArgSerSerHisSerGluArgLysSerAs 550  
QY 1470 CAATACAAAGCAACAACTGAAGCAAGAGATGACCTGCAATTCGCCCTTGGTGTACTCTGAA 1529  
Db 550 nIleArgThrArgLeuSerGlnLysLysThrHisMetLysCysProLysThrSerPheG1 570  
QY 1530 CTGCGCAACTTTTAGTTTACTC-----AAGCATCTTAACTCTGCGCATAG 1577  
Db 570 yLeuLysGlnGluHisLysValLeuIleSerLysGluLysSerSerLysAlaValHisSe 590  
QY 1578 CAGATTATCTTCAACTATGTTTATCATCCAAAAGTGCT-----AGGAT 1622  
Db 590 rAsn-----LeuHisAspIleGluAsnGlyAspGlyIleSerGluProAspTrpGlnI1 608  
QY 1623 AGATGTTTCTATCAATGAGTGTATGATGGCTCCTATGCAGAAATCCTCAGGATATTC 1682  
Db 608 eLysSerSerGlyAsnGluPheLeuSerSerLysAspGluIleHisProMetAsnLeuAl 628  
QY 1683 TCGCAACACTGGATTTGCTTTTAGTCGCAACCGACAGTTTAAGAGAACACCTATCACCA 1742  
Db 628 aGlnThrProGluGlnSerMetLysGlnAsn-----GluPheProProValSerAs 645  
QY 1743 TATTCTTGTGTCGAGCCCAACAGCAACAAACAGCATGCTGAA----- 1788  
Db 645 pLeuSerIleValGlu-----GluValSerMetGluGluSerThrGlyAspAr 661  
QY 1788 ----- 1788  
Db 661 gaspIleSerAsnAsnGlnIleLeuThrThrSerLeuArgaspLeuGlnLeuGluG1 681  
QY 1789 -----TTTCTGAATCTGAAGATGGG----- 1809  
Db 681 uLeuHisHisGlnIleProPheIleProSerGluAspSerTrpAlaValProSerGluLy 701  
QY 1810 -----GAAGTAGAACAGCAAGAACATATAGTAGTGGCCACCAATCGTCTGTA 1856  
Db 701 sAsnSerAsnLysTyrValGlnGlnGluLysGlnAsnThrAlaSerLeuSerLysVal-- 720

```
QY 1857 TTTCATAGTGTACCTGCTTACCTCTCGTCCACAGAAATGGAAGTAGTAGTGAAGA 1916
Db 721 -----AsnAlaSerArgIleLeuThrAsnAspLeuGluPheAspSerValSe 736
QY 1917 TGAAGAGTCTGTAATGCTTAAGAGAAAAAACCATTACAAAATTAAGAGTTTCTGA 1976
Db 736 rAsp-----HisSerLysThrLeuThrAsnPheSerPheGlnAlaLys 750
QY 1977 TGTTAATGAAGGAGAGAAAGATGATGAATCTCGAATCTCCATGCTCCATGATGAAGCATGG 2036
Db 750 sGlnGluSerAlaSerSerGlnThrTyrglnTyrrp---ValHisTyrLeuAspHisAs 769
QY 2037 GTTATTGTGTCACATCAATCAATCATGCTGCTGTATGCTGTTGTAGAAATATGAGACA 2096
Db 769 pSerLeuAlaAsnLysSerIleThrTyr-----GlnMetPheGlyLys 783
QY 2097 GAAATAATATAGAGAAATTTATGCGAAATCTCATGCTTCATCTAGTCAGCATCATGA 2156
Db 783 sThrLeuSerGlyThrAsnSerIleSerGlnGluIleMetAspSerValAsnAsnGluGI 803
QY 2157 CTTT-----AATCTTATTAGCATATATGTCATATAGATAAAGCTGTACCAAGCTCCGTGA 2210
Db 803 leuThrAspGluLeuLeuGlyCysLeuAlaAlaGluLeuLeuAlaLeuAspGluLysAs 823
QY 2211 AATGAGCAAAATATAGAAAAGGGGAATCTGCTCCCTCGCAACGAGAAATA----- 2265
Db 823 pAsnAsnSerCysGlnLysMetAlaAsnGluThrAspProGluAsnLeuAsnLeuValle 843
QY 2266 -----ACTGAGAAACAAAATCGGACAGCAATGATGATTAGTGAATTA 2309
Db 843 uArgTrpArgGlySerThrProLysGlu-MetGlyArgGluThrLysValLysIle- 862
QY 2310 CTCAAAAGAAAGTTTGGAAACAGATAGTGTCTCAGGGGTTTCAAACACAGACAAA 2369
Db 863 --Gln-ArgHisSerSerGly-----LeuArgIleTyrAspArgGluGluLys 877
QY 2370 -----ACAAAACCTCGAAAAGCTCTAACCC 2396
Db 878 PheLeuIleSerAsnGluLysIlePheSerGluAsnSerLeuLysSerGluGluPro 897
QY 2397 ATGTTATGACAAACACTGAAATTTACATTTTAGGGAATTCATCTCTAAGAAATATGTT 2456
Db 898 IleLeuTrpThrLysGlyGluIleLeuGlyLysGlyAlaTyrGlyThrValTyrCysGly 917
QY 2457 TTGTTT-----TTAATCATATGTTCCAAACAGGCAGCTGTAGATGAAGTAAT 2504
Db 918 LeuThrSerGlnGlyGlnLeuIleAlaValLysGlnValAlaLeuAspThrSerAsn 936

RESULT 13
US-08-118-101A-2
; Sequence 2, Application US/08118101A
; Patent No. 5620992
; GENERAL INFORMATION:
; APPLICANT: Kurtz, Stephen E.
; APPLICANT: Knickerbocker, Aron M.
; APPLICANT: McCullough, John R.
; TITLE OF INVENTION: A STRAIN OF SACCHAROMYCES CEREVISIAE
; TITLE OF INVENTION: EXPRESSING THE GENE ENCODING POTASSIUM TRANSPORTER MINK
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,101A
```

## FILLING DATE: 435

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Gaul, Timothy J.

REGISTRATION NUMBER: 33,111

REFERENCE/DOCKET NUMBER: DC27

TELECOMMUNICATION INFORMATION:

TELEPHONE: (609) 252-5901

TELEFAX: (609) 252-4526

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1235 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-118-101A-2

Alignment Scores:

Pred. No.: 4,66e-05

Score: 1235

Percent Similarity: 146.50

Best Local Similarity: 31.85%

Query Match: 20.10%

DB: 1.91%

Matches: 154

Conservative: 90

Mismatches: 227

Indels: 295

Gaps: 39

US-09-874-162A-7 (1-4409) x US-08-118-101A-2 (1-1235)

QY 451 TATAGATTCTTCGAACTCGGAATCTCATAGCAACCAATATTTTGCACAGAACTTACT 510

Db 148 PheLysMetArgThrLysThrIleLeuGluArgGluLeuThrAlaArgThrMetThr 167

QY 511 TACATGTCTCATCGAAACTCCAGCAACCAACATCAAAAGGAAAAACATTTAAAGTTGATGAT 570

Db 168 LysAsnArgThrGlyThrGlnArgThrSerTyrProArgLysGlnAlaLysThrAspAsp 187

QY 571 ATGTTATCAAAAGTA-----GAGAAAATGAAA 597

Db 188 PheGlnGluLysLeuPheSerGlyGluMetValAsnArgAspGluGlnAspSerValHis 207

QY 598 GGAGAGCAAAATCTCATAGCTTGTGTCAGCTCAATTCGAGCTTACGTTTACTGCTTTC-- 654

Db 208 SerAspGlnAsnSerHisAspIleSerArgAspSerSerAsnAsnAsnThrAsnHisAsn 227

QY 655 -----TTCCACAAAATGATAAGCCATCATCCAAAACCTCAGAA 690

Db 228 GlySerSerGlySerLeuAspAspPheValLysGluAspGluThrAspAspAsnGlyGlu 247

QY 691 AATGAAACAAAATCTGTTACCTCGGAAGCTCTGTTGAAAGTTGCCACACAAAAAGA 750

Db 248 TyrGlnGluAsnAsn----- 252

QY 751 AAGGATGTTAAGTTGTCCAATAAGGCAAGTTCCACAGTAAAAAGCGGTGCCTTTGATT 810

Db 252 ----- 252

QY 811 CTTGACCTCAATCAAAACAAACCCGGAATTTCCCGTCCCTGCTTCCAGTTTCCAGTAATGAA 870

Db 253 -----SerTyrSerThrValGlySerSerSerAsnThr 263

QY 871 TTT---GAACCTAGTAACGCCATATGTTGTAAGCTCTTACTCGTTGCTATT----- 918

Db 264 ValAlaAspGluSerLeuAsnGlnLysProLysProSerSerLeuArgPheAspGluPro 283

QY 919 ---AGATGACTCGTCCAGAGAGAGAGAGTTAATGGAATGATTAAATGGAGAAACCAAT 975

Db 284 HisSerLysGlnArgProAlaArg----- 291

QY 976 GAAATATTGATGTCATGAGAGCTTCCAGCCAGAGAAACGAAATCGTGAGATGGG 1035

Db 292 -----ValProSerGluLysPheAlaLysArgGlySerArgAsp----- 305

QY 1036 GAAAAGACATTTGTTGCAAAATGATGATTGTAATAAAACAGCGCTTACAGCTTTTA 1095

Db 306 -----IleSerProAlaAspMetTyrArgSerIleMetMetLeu 318  
 QY 1096 GATGGGAATATGAAGTAGCATCGAAGAAATGAAGATGTCATA----- 1143  
 Db 319 GlnGlyLysHisGlu---AlaThrAlaGluAspGluGlyProProLeuValIleGlySer 337  
 QY 1144 -----AGCAAGAAAGAGCAACATGGGAG 1167  
 Db 338 ProAlaAspGlyThrArgTyrLysSerAsnValAsnLysLeuLysLysAlaThrGly--- 356  
 QY 1168 ACTATTCCTGTGAGGAAGAGCTCCCTCCATTCGAACATTTCTCAGGACCTACGTTG 1227  
 Db 357 -----IleAsnGlyAsnLysIleArgSpLysGlyAsnGluSerAsnThrAsp 374  
 QY 1228 CAGTTCACCTCTGTTGGACGAGGACCAATCAATAAATCTACGGCTCCTATTGCCAAA 1287  
 Db 375 GlnAsnSerVal-----SerSerGlnAlaAsn-----SerThrAlaSerValSer--- 389  
 QY 1288 CCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAA-----AACAAGCCTGGTTCA 1341  
 Db 390 -----AspGluSerSerLeuHisThrAsnPheGlyAsnLysValProSer 404  
 QY 1342 GTTAAACCTACTCAAACT-----ATTGCTGTTAAAGAAATCATTTG 1380  
 Db 405 LeuArg---ThrAsnThrHisArgSerAsnSerGlyProIleAlaIleThrAspAsnAla 423  
 QY 1381 ACTACAGATCTACAAACAAGAAAGAAAGGATACTCCAAATGAAACCCGACAAATA 1440  
 Db 424 GluThrAsp-----LysLysHisGlyProSer----- 432  
 QY 1441 AGAATATTTTATGTTTCTCTATAACAACAATCAAGG-----CAACAACCTGAA 1491  
 Db 433 -----IleGlnPheAspIleThrLysProProArgLysIleSerLysArgValSer 449  
 QY 1492 GCAAGAGATGACCTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1548  
 Db 450 ThrPheAspAspLeuAsn---ProLysSerValLeuTyrArgLysLysAlaSerLys 468  
 QY 1549 -----TTACTCAAGCATCTTAAACTCTGCCATACAGAGATTATCTTCAACTATGTTAT 1602  
 Db 469 LysTyrLeuMetLysHisPhe----- 475  
 QY 1603 CATCCAAAGTGCTAGATAGATGTTCTTCTATC---RATGAGTGTATGATGCTCTCTAT 1659  
 Db 476 ---ProLysAlaArgArgIleArgGlnGlnIleLysArgArgLeuSerThrGlySerIle 494  
 QY 1660 GCAGGAATCCTCAGGATATTCTATCGCCCAACCTGGATTGCTTTAGTCGCAACGGACCA 1719  
 Db 495 GluLysAsn-----SerSerAsnAsnVal 502  
 QY 1720 GTTAAGAGAACCTTATCACACATATCTTGTGTGCGAGCCCAACCAAGCAAGCAAGC 1779  
 Db 503 SerAspArgLysProIleThr----- 510  
 QY 1780 ATGCTCTGAA-----TTTCTT 1794  
 Db 511 MetAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 530  
 QY 1795 GAATCTGAAGATGGGGAAGTAGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGC 1854  
 Db 531 AspAsnGluSerGlyAsp---GluAspGluArgValGlnGlnSerGluPro----- 546  
 QY 1855 TATTTCCATGATGATACCTGCTTACCTCTCCGTCACAGAAATGAA----- 1902  
 Db 547 -----HisSerAspSerGluLeuLysSerHisGlnGlnGlnGlnGlnLysHisGlnLeu 564  
 QY 1903 -----GTAGATAGTGAAGATGAAAGGAT 1926  
 Db 565 GlnGlnAsnLeuHisArgMetTyrLysThrLysSerPheAspAspAsnArgSerArgAla 584  
 QY 1927 CTGNATCGCTAAGAGAAAACCTTACAAATTTAGAGTTTCTGATGTTAATGAA 1986  
 Db 585 ValProMetGluArgSerArgThrIle---AspMetAlaGluAlaLysAspLeuAsnGlu 603

QY 1987 -----GGAGAGAAAGAAAGTGTATGAACCTCTGCAATCTCCAT----- 2022  
 Db 604 LeuAlaArgThrProAspPheGlnLysMetValTyrGlnAsnThrLysAlaHisHisArg 623  
 QY 2023 -----GTCATGAAGCATGGG---TTT 2040  
 Db 624 LysLysProAsnPheArgLysArgGlyTrpAsnAsnLysIlePheGluHisGlyProTyr 643  
 QY 2041 ATTGCTGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2100  
 Db 644 AlaSerAspSerAspArgAsnTyrProAspAsn-----SerAsnThrGly----- 658  
 QY 2101 ATAATTAAGAAGAAATTTATGTCGAACCTTCATGCTTCTCATCTAGTCAGCATGCATGCTTT 2160  
 Db 659 -----AsnSerIleLeuHisTyrAlaGlu----- 666  
 QY 2161 AATCTTATGATATATGTCATATAGATAAAGCTGTACCAAGCTCCGTGAATGCAGCAA 2220  
 Db 667 -----SerIleLeuHisAspGlySer-----His 675  
 QY 2221 AAATTTAGAAAAGGGCAATCTCTCCCTCGCAACGAAAGAAATAAATCAAGAAACAAAAT 2280  
 Db 676 LysAsnGlySerGluGluAlaSerSerAspSerAsnGluAsnIleTyrSerThrAsnGly 695  
 QY 2281 GGGACA-----GCAAAATGGATTACTGAAATTAATCACTAAAAGAAAGCT--- 2325  
 Db 696 GlySerAspHisAsnGlyLeuAsnAsnTyrProThrTyrAsnAspAspGluGluGlyTyr 715  
 QY 2326 -----TTGGAACACAGATAGTCTCAGGGGTTTCAAAACAG 2361  
 Db 716 TyrGlyLeuHisPheAspThrAspTyrAspLeuAspProArgHisAspLeuSerLysGly 735  
 QY 2362 AGCAAAAACAAAAACTC 2379  
 Db 736 SerGlyLysThrTyrLeu 741  
 RESULT 14  
 US-09-418-710-21  
 ; Sequence 21, Application US/09418710  
 ; Patent No. 6596482  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jones, Michael H.  
 ; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR  
 ; FILE REFERENCE: 06501-042001  
 ; CURRENT APPLICATION NUMBER: US/09/418,710  
 ; PRIOR FILING DATE: 1999-10-15  
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01783  
 ; PRIOR FILING DATE: 1998-04-17  
 ; PRIOR APPLICATION NUMBER: JP 9/310027  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: JP 9/116570  
 ; PRIOR FILING DATE: 1997-04-18  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 21  
 ; LENGTH: 1972  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 US-09-418-710-21  
 Alignment Scores:  
 Pred. No.: 6,54e-05 Length: 1972  
 Score: 146.00 Matches: 154  
 Percent Similarity: 32.90% Conservative: 126  
 Best Local Similarity: 18.10% Mismatches: 314  
 Query Match: 1.90% Indels: 257  
 DB: 4 Gaps: 33  
 US-09-874-162A-7 (1-4409) x US-09-418-710-21 (1-1972)  
 QY 139 ATCGAGCATCGAGACCAACCATCGATCCAGATCCACGGTTTATAGAAAAACAGAA 198

Db 83 IleGluSerGluAspSerAspSerGluSerGluAlaGlnHisLysSerAsn 102  
QY 199 TTAAGAGCAACCTATGTCCTGAGTTACATAAATAGATTCATGACAGATGCTGCC 258  
Db 103 AsnGln-----ValLeuLeuHisGlyIleSerAspProLysAlaAspGlyGln 118  
QY 259 CGCCGAGAGAGAGTCCCTAAGAGAGAGATTCAGCCGAAGCTCTCGTGACTCTGTC 318  
Db 119 LysAlaThrGluLysAlaGlnGluLysArgIleHisGlnProLeuProLeuAlaPheGlu 138  
QY 319 AGTCAGTGTCTCGAGGGAATGTCTCACTCCGCCACGCACAGC-----AGTGAAGC 372  
Db 139 SerGlnThr-----HisSerPheGlnSerGlnGln 148  
QY 373 CTACTCCCGCCGTC-----ACCCACCCATCACCCCTCTCTTCATTCGCG 420  
Db 149 LysGlnProGlnValLeuSerGlnGlnLeuProPheIlePheGlnSerSerGlnAlaLys 168  
QY 421 AGCAGCACTCCGACAGCAACACAGATCTATAGATTCTTCGAACCTCGGATCTCAT 480  
Db 169 GluGluSerValAsnLysHisThrSerValIleGlnSerThrGlyLeuValSerAsnVal 188  
QY 481 GCACCAATATTTTG-----CACAGACTCTTACTTACATGCTCT----- 519  
Db 189 LysProLeuSerLeuValAsnGlnAlaLysLysGluThrTyrMetLysLeuIleValPro 208  
QY 520 -----CATCGAACTCCAGAACCAAAACATCAAAAGGAAACATTTAAAGTT 564  
Db 209 SerProAspValLeuLysAlaGlyAsnLysAsnThrSerGluGluSerSerLeuLeuThr 228  
QY 565 GATGATATGTTATCAAAAGTAGAG-----AAATGAAGGA 600  
Db 229 SerGluLeuArgSerLysArgGluGlnTyrLysGlnAlaPheProSerGlnLysLys 248  
QY 601 GAGCAAGATCTCATAGTTCTCA----- 624  
Db 249 GlnGluSerSerLysSerLysLysLysValIleAlaLeuSerAsnProLysAlaThr 268  
QY 625 -----GCTCATTTGAGCTTACGTTTACTGTTTCTTCCACAAAATGATAAG 672  
Db 269 SerSerProAlaHisProLysGlnThrLeu-----GluAsnAsnHis 283  
QY 673 CCATCACCAACTCGAANAAT-----GACAAAATCTGTTACCTG 714  
Db 284 ProAsnProPheLeuThrAsnAlaLeuLeuGlyAsnHisGlnProAsnGlyValIleGln 303  
QY 715 GAAGTCTCTG-----CTGTGAAGTTTGGCAGCAAAA 747  
Db 304 SerValIleGlnAlaProLeuAlaLeuThrLysThrLysMetGlnSerLysIle 323  
QY 748 AGAAGGATGTA-----AGTTGTCCAATAAGGCAAGTTCCC 783  
Db 324 AsnGluAsnIleAlaAlaSerSerThrProPheSerSerProValAsnLeuSerThr 343  
QY 784 ACAGTAAAG-----CAGTGCTTTGATCTCTGACCTCAATCAACAAA 831  
Db 344 SerGlyArgArgThrProGlyAsnGlnThrProValMetProSerAlaSerProIleLeu 363  
QY 832 CCGGAAATTTCCCTCCCTTCAGTTTCCAGTAATGAATTTGAACCT----- 879  
Db 364 HisSerGlnGlyLysGluLysAlaValSerAsnAsn-----ValAsnProValLysThrGln 382  
QY 880 AGTAACAGCATATGCTGAGCTTACTGCTGTTGCTATTAGAGTACCTCGT----- 930  
Db 383 HisHisSerHisProAlaLysSerLeuValGluGlnPheArgGlyThrAspSerAspIle 402  
QY 931 CCAGGAAGAAGAGATTTAATGAATGATTATGAGAGAACCAATGAAATATTGATGTC 990  
Db 403 ProSerSerLysAsp-----SerGluAspSerAsnGluAspGluGlu 417  
QY 991 AATGAAGACTTCCAGCCAGAGAAACCAATCGTGAGGATGGGGAAGACATCTGTT 1050  
Db 418 AspAspGlu-----GluGluAspGluGluAspGlu----- 428

QY 1051 GCACAAATGACAGTATTGATAAAACAGCGCTTACAGCTTTTGTAGTGGGAATATGAA 1110  
Db 429 -----AspAspGluSerAsp 433  
QY 1111 GTAGCCATGCAGGAAATGGAAGATGTCCTAATAGCAAGAAAGAGCAACATGGGAGACT 1170  
Db 434 AspSerGlnSerGluSerAspSerAsnSerGluSerAspThrGluGlySerGluGluGlu 453  
QY 1171 ATTCTTGATGGGAAGGCTGCTCCATTCGAAATTTTCTCAGGACCTTACGTTGCAG 1230  
Db 454 AspAspAspLysAspGlnAspGluSerAspSerAspThrGluGlyGluLysThrSer 473  
QY 1231 TTCATCTTCTGTCGACAGAGAGACCAATGATAAATCTAGCGCTCTTATTCGCAACCT 1290  
Db 474 MetLysLeu-----AsnLysThrThrSerSerValLysSerPro 486  
QY 1291 CTTGCCACTAGAAATTCAGAGAGT-----CTCCATCAGGAAAACAGCCCTGGT 1338  
Db 487 SerMetSerLeuThrGlyHisSerThrProArgAsnLeuHisIleAlaLysAlaProGly 506  
QY 1339 TCAGTT-----AAACCTACTCAAACTATTGCTGTATAAGAAATCATTTGACT 1383  
Db 507 SerAlaProAlaAlaLeuCysSerGluSerGlnSerProAlaPheLeuGlyThrSerSer 526  
QY 1384 ACAGATCTACAACAAGAAAGAAAGGATCTCCAAATGAAACCGA----- 1431  
Db 527 SerThrLeuThrSerSerProHisSerGlyThrSerLysArgArgValThrAspGlu 546  
QY 1432 CAATAATTAAGAAATTTTATCAGTTTCTCTATAACAACAATACAAG----- 1479  
Db 547 ArgGluLeuArgIleProLeuGluTyrGlyTyrGlnArgGluThrArgIleAsnAsnPhe 566  
QY 1480 ---CAACAACACTGAAGCAAGAGATGACCTGCCTGCTGCTGCTGCTGCTGCTGCTG 1536  
Db 567 GlyGlyArgLeuGlnGlyGluValAlaTyrTyAlaPro---CysGlyLysLysLeuArg 585  
QY 1537 AAATTTTATAGTTTACTCAAGCATCTTAAACTCTGCCATAGCAGATTATCTTCAACTAT 1596  
Db 586 GlnTyrProGluValIleLysTyrLeuSer----- 595  
QY 1597 GTTTATCATCAAAGGTGCTAGGATAGATGTTTCTTATCAATCAGTGTATTATGATGGCTCC 1656  
Db 596 -----ArgAsnGlyIleMetAspIleSer----- 603  
QY 1657 TATGCGAATCCTCAGGATATTCATCGCCCAACCTGGATTGCTTTTACTCCGACGGA 1716  
Db 604 -----ArgAspAsnPheSer----- 610  
QY 1717 CCAGTTAAGAGAACACCTATCACACATATTCTTGTGTGAGGCCAAAACGAAAGCA 1776  
Db 611 -----AlaLysIle 613  
QY 1777 AGCATGCTCAATTTCTTGAATCTGAAGATGGGAGATGAGCAAGCAAGAACATATAGT 1836  
Db 614 ArgValGlyAspPheTyrGluAlaArgAspGlyProGlnGluMetGlnTyrCys----- 631  
QY 1837 AGTGGCCACAATCGTCTGTTATTCATAGTAGTACCTGCTTACCTCTCCGTCACCAAGAA 1896  
Db 632 -----LeuLeuLysGluLysValIleProArgIleArgAlaMetGlu 646  
QY 1897 ATGGAAGTAGATAGTGAAGATGAAAGGATCTGGAATGGCTAAGAGAAAACCATTTACA 1956  
Db 647 GlyArgArgGlyArgProProAsnProAspArgGlnArgAlaArgGluGluSerArgMet 666  
QY 1957 CAAATGAACAGTTTCTGATGTTTAAAGGAGAGAAAGTGTGAACTCTGGAAT 2016  
Db 667 ArgArgArgLysGlyArgProProAsnValGlyAsnAlaGlu----- 680  
QY 2017 CTCCATGTCTAAGACATGGGTTTATTGTGTGACAAATCAATGAATCATGCTGTATGCTG 2076  
Db 680 ----- 680



QY 1687 CAACCTGGATTTGCTTTTATGTCGCAACGACCGAGTAAAGAACACACCTATCATCATATT 1746  
Db 398 -----ArgThrAspLeuProAsnSer 404  
QY 1747 CTTGTGTGCGAGGCCAAACGACAAAGCAAGCATGCTGTAATTTCTTGAATCTCAAGAT 1806  
Db 405 ValVal-----AsnAspAsp 409  
QY 1807 GGGGAAGTAGAACAGCAA-----AGAACATATAGTAGTGCCACAATCGTCTGTAT 1857  
Db 410 IleGluTyrAsnGlnLysSerAspAsnAsnSerTyrSerThr-----AsnAsnLeuTyr 427  
QY 1858 TTCCATAGTGATACCTGCTTACTCTCCGTCCACAGAATGGAAGTAGTAGAAGAT 1917  
Db 427 ----- 427  
QY 1918 GAAAGGATCTCTGAATGCTAAGAGAAAAAACCATTACAAAATGAAGAGTTTCTGAT 1977  
Db 428 -----AsnAsnIleAsnMetThrGlnAsnGlnAspAsnAsnAsn 440  
QY 1978 GTTAATGAAGGAGAGAAAGAGTGAATAACTCTGGAATCTCCATGTCATGAAGCATGGG 2037  
Db 441 ValAsn-----IlePheLys-----HisValGlnAsnAspSer 451  
QY 2038 TTTATTGCTGACAATCAAATGAATCATGCTGATGCTGTTGTAGAA----- 2085  
Db 452 PheGlnCysPheAsnSerAsnAsn-----LeuTyrIleGluLysAspIleLys 467  
QY 2086 -----AATTTGGACAGAAAAATAATTAGAAAGATTTATGT-----CGAACTTCATG 2133  
Db 468 GluAsnAsnIleSerGln-----IleAsnArgLysLeuLysCysThrLysArgAsnPhe--- 484  
QY 2134 CTTTCATCTAGTCAGATGATGACTTTAATCTATTAGCATTAATGCTCAATAGATAAAGCT 2193  
Db 485 -----ThrLysLysSerArgLysLysIleAsnThrValThrTyrLeuGlnIleAspLys--- 501  
QY 2194 GTTACCAAGCTCGTGAATGAGCAAAATAATAGAAAAGGGGGAATCTGCTTCCCTGCA 2253  
Db 502 ValIleLysIle-LeuLysCysLys----- 509  
QY 2254 AACGAGAAATAACTCAGAACAC-----AAATGGACAGCAAAATGATTTAGTGAAT 2307  
Db 510 -----LysLysTyrIleLysHisIleLysLysMetLysTyrMetAsnAsnPheGlnAsnPh 528  
QY 2308 AACTCAAAAGAGAAAGCTTTGGAACACAGATAGTCTCAGGGGTTTCAAAACAGAGCA-- 2365  
Db 528 elysLysLeuLysLysLeu-----GlnLysPheGlnAsnAlaSerPh 542  
QY 2366 -----AAAAACAARAACTCTGAAAAGCTCTAACCCCATGTTATGGA 2406  
Db 542 eglLeuLysIleAsnLysIleAsnLysAsnIleArgArgLeu-----AsnLysLeuLys 560  
QY 2407 CAACACTGAATTAATTTAGGGAATTCATCC-----TCTAA 2445  
Db 560 slsArgLysAsnHisSerIleAsnIleProValThrSerIleGluTrpLeuAsnAs 580  
QY 2446 GAATTAATGTTTGTGTTT-----AATCATATGTTCCAAACAGCAGCACTGTAGTGAAGTAA 2502  
Db 580 nSerTyrThrPheAspPheIleAsnAsnSerIleGlnSer----- 593  
QY 2503 ATGATTTCAACAGGATATTGTCAGGGT-----TCTACTTCATTCATTATGC 2553  
Db 594 -----ThrSerTyrProTrpLysAsnLysCysAspAlaThrIleArgAsnHisLe 610  
QY 2554 AGCATTACATGATATCACTTTTATTGATGTCATTAAACATCTCTGTACTTTAAGCATGA 2613  
Db 610 uHisLeuHis-----AsnValIleIleAspLysAsnAs 621  
QY 2614 AAGCAATATTTCAAGATATTTTAACTCAACAATGTCATCAA----- 2659  
Db 621 nLysThrTyrPheMetLysAsnLeuValGluAsnArgIleValArgAsnIleIleSerLy 641  
QY 2660 -----TATGTTGAATGATCTAGAAATTTATTCATATATAAT 2697

Db 641 sglLysLysCysglNserLeuTyr-LysAsnLysGlnAsnValTyrPheCysTyrLysA 661  
QY 2698 CAGAAATTTTTCATTTATGACGGCTGTTTTTCTACTTTTGAATTTGTGAGACATTTTC 2757  
Db 661 snAsnPheSerLeuLeuLysSerSerIle----- 670  
QY 2758 TTGGGAGGGGAAAAATTTGGAATGTTCCCTTTTGTAGAAATTTGAAGTGGTCTTCATATGTC 2817  
Db 671 -----LeuLysPheIleCysC 676  
QY 2818 AACTACAGAAAGAAAAAATA-----GAAATTTGAAGATTTTATGAAATTTATATTG 2871  
Db 676 ysIleLysThrLeuLysLysMetPheAsnAlaPheThrAsnSerThrTyrAsnThrLysP 696  
QY 2872 CATTACTATTGCACTCAAACTTTTGATCCTTTGTTTGAATCATTTGTCAATTCGGAAT 2931  
Db 696 helIleLeuPheLeuIleSer-----TyrMet-AsnLysMetLeuIleLysAsn 711  
QY 2932 GAAAAATATAATGTAATTTTACATTTACATTAAGTTCCTTTTACAA----- 2976  
Db 712 LysLysLeu-----LysPhe--ValLysLeuPheLeuIleGlnThrAlaIleGluAla 728  
QY 2977 -----TTAAAAAATAGCACTTCTTCATCTTATGCTGCTTTTGAAGATAT 3021  
Db 729 PheArgTyrAlaArgIlePheAsnGlnAspSerPheTyrProCysLeuGlnHisPhe 748  
QY 3022 ---TAAATTTTCATTTGTTGACATGAAATGCTATGTTTGTGTTTATAAGATTACAGACCA 3078  
Db 749 ArgLysIleLysLysArgLeuIleAsnLysTyrLysIleGly----- 762  
QY 3079 TTGTTTTTCATGTCGATAATTTTAGTCATGCTCACCCCGTATGTTTTTTTTTTTAAAC 3138  
Db 763 -----HisAsn 764  
QY 3139 TTGAACATTTGCTGTTTGTGTTTCTTTTCTTTTAAAT 3174  
Db 765 LysAsnLeuLeuArgGluPhePheLeuPheAsn 776

Search completed: August 25, 2004, 19:07:37  
Job time : 205 secs





Alignment scores:  
Pred. No.: 0      Length: 776

Score:	4078.00	Matches:	776
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	53.03%	Indels:	0
	9	Gaps:	0

US-09-874-162A-7 (1-4409) x US-09-874-162A-8 (1-776)

Qy	52	ATGACAGCATCGCGCGCGCTCTTCTTCTCCAAATCACTCCGCAATTCGGGGCTGCGGA	111
Db	1	MetThrGlyIleAlaAlaAsPhePheSerAsnThrCysArgPheGlyGlyCysGly	20
Qy	112	CTCCACTTCCCAACCCCTGGCGAGCTCATCGAGCACATCGAGGACAACCAATCATGATACA	171
Db	21	LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr	40
Qy	172	GATCCACGGGTTTATAGAAAAACAAGAAATTACAGAGCCAACTATGTTGCCTGAGTTAC	231
Db	41	AspProArgValLeuGluLysGlnGluLeuGlnProThrTyrValAlaLeuSerTyr	60
Qy	232	ATAAATAGATTATGACAGATGCTCCGCGCAGAGCAGGAGTCCCTAAAGAAAGAAGATT	291
Db	61	IleAsnArgPheMetThrAspAlaIleArgArgGluGlnGluSerLeuLysLysLysIle	80
Qy	292	CAGCGAAGCTCTCGTGACTCTCTCGAGTCACTGTCTCGAGGGAATGTGTCACATCCC	351
Db	81	GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro	100
Qy	352	CCAGCCACAGCAGTGGGAAGCCTTACTCCGCCGCTGACCCACCCATCACCCCTCTCT	411
Db	101	ProArgHisSerSerGlySerLeuThrProProValThrProProlIleThrProSerSer	120
Qy	412	TCATTCGCCAGCAGCTCCGACAGAGCCAAACACAGATCTATAGATTCTTCGAACTCGG	471
Db	121	SerPheArgSerSerThrProThrGluProThrGlnIleTyArgPheLeuArgThrArg	140
Qy	472	AATCTCATAGACCAATATTTTCCACAGAACTCTTACTTACATGTCTCATCGAAATCCC	531
Db	141	AsnLeuIleAlaProlIlePheLeuHisArgThrLeuThrTyMetSerHisArgAsnSer	160
Qy	532	AGAAACAACATCAAAAGGAAACATTAAAGTTGATGATATGTTATCAAAAGTAGACAAA	591
Db	161	ArgThrAsnIleLysArgLysThrPheLysValAspAspMetLeuSerLysValGluLys	180
Qy	592	ATGAAAGCAGAGCAAGATCTCATAGCTGTGCAGCTCATTTGCAGCTTAGCTTACTGGT	651
Db	181	MetLysGlyGluGlnGluSerHisSerLeuSerAlaHisLeuGlnLeuThrPheThrGly	200
Qy	652	TTCTTCCACAAAAATGATAAGCCATCACCAACCTCAGAAAATGACAAAAATTCGTGTACC	711
Db	201	PhePheHisLysAsnAspLysProSerProAsnSerGluAsnGluGlnAsnSerValThr	220
Qy	712	CTGGAAGTCTGCTGTCAAGTTTCCACAAAAAAGAAAGATGAAGTTGTCCCAATA	771
Db	221	LeuGluValLeuLeuValLysValCysHisLysLysArgLysAspValSerCysProIle	240
Qy	772	AGGCAAGTTCACAGGTAAAGAACAGAGTGCCTTTGATTCTCGACCTCAATCAAAACAAA	831
Db	241	ArgGlnValProThrGlyLysLysGlnValProLeuIleProAspLeuAsnGlnThrLys	260
Qy	832	CCCGGAATTTCCGTCCTTGCAGTTTCCAGTAATGAATTGAACTAGTAGACAGCAT	891
Db	261	ProGlyAsnPheProSerLeuAlaValSerSerAsnGluPheGluProSerAsnSerHis	280
Qy	892	ATGGTGAAGTCTTACTCGTTGCTATTAGAGTGACTCGTCAGGAAGAAAGAGCTTTAAT	951
Db	281	MetValLysSerTyrSerLeuLeuPheArgValThrArgProGlyArgGluPheAsn	300
Qy	952	GGAATGATTATGAGAGAAACCAATGAAAATTTGATGTCATGAGAGCTTCCAGCCAGA	1011
Db	301	GlyMetIleAsnGlyGluThrAsnGluAsnIleAspValAsnGluGluLeuProAlaArg	320
Qy	1012	AGAAAACCAATCGTAGAGTAGGGGAAAAAGACATTTCGCACAAATGACAGATTATTGAT	1071

[illegible]



```

QY 1876 TTACCTCTCCGTCACAAAGAAATGAAGTAGATAGTGAAGATGAAAGGATCCTGAATGG 1935
Db 572 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluAspProGluTrp 591
QY 1936 CTAAGAGAAAAACCATACAAATGAAGAGTTTCTGATGTAATGAAGAGAGAGAAA 1995
Db 592 LeuArgGluYsThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluYs 611
QY 1996 GAAGTGAATCACTCTGAATCTCCATGTCATGAAGCATGGTTTATTGCTGACAATCAA 2055
Db 612 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 631
QY 2056 ATGAATCATGCTGTATGCTGTTTCTAGAAAATTAAGCAGAGAAAATTAAGAAGAA 2115
Db 632 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsn 651
QY 2116 TTATGTCGAACTTCATGCTTCATCTAGTCAGCATGCGATGCTTAATCTTATTAGCATA 2175
Db 652 LeuCysArgAsnPheMetLeuHisValSerMetHisAspPheAsnLeuIleSerIle 671
QY 2176 ATGTCATAGATAAAGCTGTTACCAAGCTCCGTGAAATGCAGCAAAATTAAGAAAAGGG 2235
Db 672 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluYsGly 691
QY 2236 GAATCTGCTCCCTGCAAAACGAGAAATTAAGTGAAGCAAAATGGGACAGCAAAATGGA 2295
Db 692 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 711
QY 2296 TTTAGTGAATTAATCTCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGGTTTCA 2355
Db 712 PheSerGluIleAsnSerLysGluYsAlaLeuGluThrAspSerValSerGlyValSer 731
QY 2356 AAACAGAGCAAAAAACAATACTC 2379
Db 732 LysGlnSerLysLysGlnLysLeu 739

RESULT 3
US-09-764-864-931
; Sequence 931, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 931
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-931

Alignment Scores:
Pred. No.: 1,8e-169 Length: 388
Score: 2043.00 Matches: 386
Percent Similarity: 99.48% Conservations: 0
Best Local Similarity: 99.48% Mismatches: 2
Query Match: 26.57% Indels: 0
DB: 9 Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-931 (1-388)

QY 1216 GGACCTAGCTTCCAGTTTCACTCTCGTTGGACAGAGACCAATGATAAATCTACGGCT 1275
Db 1 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla 20
QY 1276 CCTATTGCCAAACCTCTTCCCATAGAAATTCAGAGAGTCTCCATCAGGAAACCAAGCCT 1335
Db 21 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 40

```

```

QY 1336 GGTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAATCATTGACTACAGATCTACAA 1395
Db 41 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGln 60
QY 1396 ACAAGAAAAAGAAAGATCTCCAAATGAAGAAAAACCCGACAAAAATTAAGAAATATTTATCAG 1455
Db 61 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheYrGln 80
QY 1456 TTTCTCTATAACAACAATCAAGGCAACAACTGAAGCAAGAGATGACCTGATTCGCTC 1515
Db 81 PheLeuYrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysPro 100
QY 1516 TGTGTACTCTGAACCTGCCGCAAACTTTTATAGTTTACTCAACATCTTAAACTCTGCCAT 1575
Db 101 TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHis 120
QY 1576 AGCAGATTTTATCTTCAACTATCTTTATCATCCAAAAGGTGCTAGGATAGATGTTTCTATC 1635
Db 121 SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle 140
QY 1636 AATGAGTGTATGATCGCTCTCTATGAGGAAATCTCTCAGATATTTCATCGCCAACTGGA 1695
Db 141 AsnGluGlyTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly 160
QY 1696 TTTGCTTTTGTAGTCGCAACGACAGTAAAGAGAACACCTATCACACATATTTCTGTGTC 1755
Db 161 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 180
QY 1756 AGSCCAAAACGACAAACCAAGCATGCTGAAATTTCTGAACTGGAAGATGGGCAAGTA 1815
Db 181 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 200
QY 1816 GAACAGCAAAAGAACATATAGTAGTGGCCACAACTCTGTATTTCCATAGTATACCTGTC 1875
Db 201 GluGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisArgAspThrCys 220
QY 1876 TTACCTCTCCGTCACAAAGAAATGGAAGTAGTGAAGATGAAAGAGATCCCAATGG 1935
Db 221 LeuProLeuArgProGlnGlnMetGluValAspSerGluAspGluLysAspProGluTrp 240
QY 1936 CTAAGAGAAAAAACCATACAAATTAAGAGAGTTTCTGATGTTAATGAAGAGAGAGAAA 1995
Db 241 LeuArgGluYsThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluYs 260
QY 1996 GAAGTGAATGAAACTCTGGAATCTCCATGTCATGAAGCATGGTTTATTGCTGACAAATCAA 2055
Db 261 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 280
QY 2056 ATGAATCATGCTGTATGCTGTTTCTAGAAAATTAAGCAGAGAAAATTAAGAAGAAAT 2115
Db 281 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLysLysAsn 300
QY 2116 TTATGTCGAACTTCATGCTTCATCTAGTCAGCATGATGCTTTTAAATCTTATTAGCATA 2175
Db 301 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 320
QY 2176 ATGTCATAGATAAAGCTGTTACCAAGCTCCGTGAAATGCAGCAAAATTAAGAAGGGG 2235
Db 321 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluYsGly 340
QY 2236 GAATCTGCTCCCTGCAAAACGAGAAATTAAGTGAAGCAAAATTAAGGACAGCAAAATGGA 2295
Db 341 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 360
QY 2296 TTTAGTGAATTAATCTCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGGTTTCA 2355
Db 361 PheSerGluIleAsnSerLysLysAlaLeuGluThrAspSerValSerGlyValSer 380
QY 2356 AAACAGAGCAAAAAACAATACTC 2379
Db 381 LysGlnSerLysLysGlnLysLeu 388

```

US-09-874-162A-911  
; Sequence 911, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 911  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-764-864-911

Alignment Scores:  
Pred. No.: 9, 43e-122 Length: 289  
Score: 1497.00 Matches: 283  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.65% Mismatches: 0  
Query Match: 19.47% Indels: 0  
DB: 9 Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-911 (1-289)

QY 1528 AACTGCGCAAACTTTATAGTTTACTCAAGCATCTTAACCTCTGCCATAGCAGATTATC 1587  
:::|||||  
Db 6 AspCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle 25

QY 1588 TTCAACTATGTTTATCATCCAAAAGTGCTAGATAGATGTTTCTATCAATGAGTGTAT 1647  
|||||  
Db 26 PheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIleAsnGluCysTyr 45

QY 1648 GATGGCTCTATGCGAGAAATCCTCAGGATATATCATGCGCACTCGGATTTGCTTTAGT 1707  
|||||  
Db 46 AspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGlyPheAlaPheSer 65

QY 1708 CGCAACGGACAGTAAAGAGAACACTATCACACATATCTTGTCGAGGCCAAAACGA 1767  
|||||  
Db 66 ArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCysArgProLysArg 85

QY 1768 ACAAAGCAAGCATGCTGAATTTCTTGAACTCGAAGATGGGGAAGTAGACACAGCAAGA 1827  
|||||  
Db 86 ThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluValGluGlnArg 105

QY 1828 ACATATAGTAGTGGCCACATCGTCTGATTTCCATAGTGATCTGCTTACCTCTCGGT 1887  
|||||  
Db 106 ThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerAspThrCysLeuProLeuArg 125

QY 1888 CCACAGAAATCGAAGTAGATAGTGAAGTAAAGAGGATCTCGAATGCTTAAGAGAAAAA 1947  
|||||  
Db 126 ProGlnGluMetGluValAspSerGluAspGluLysAspProGluTyrLeuArgGluLys 145

QY 1948 ACCATTACAAATTAAGAGAGTTTCTGATGTTATGAAGAGAGAGAAAGATGATGAAA 2007  
|||||  
Db 146 ThrIleThrGlnIleGluPheSerAspValAsnGluGlyGluLysGluValMetLys 165

QY 2008 CTCTGGAATCTCCATGTCATGAAGCATGGGTTTATGTGTCACATCAATCAATCAATGCC 2067  
|||||  
Db 166 LeuTyrAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGlnMetAsnHisAla 185

QY 2068 TGTATGCTGTTGTAGAAAATATGGACAGAAAATAATTAAGAGAAATTTATGCGAAAC 2127  
|||||  
Db 186 CysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsnLeuCysArgAsn 205

QY 2128 TTCATGCTTCATCTAGTCAGCATGATCTTATCTTATTTAGCATATGCTCAATAGAT 2187  
|||||  
Db 206 PheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIleMetSerIleAsp 225

QY 2188 AAAGCTGTTACCAAGCTCCGTGAAATGCAGCAAAAATAATAGAAAAAGGGGGAATCTGCTTCC 2247  
|||||

Db 226 LysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGlyGluSerAlaSer 245  
|||||

QY 2248 CCTGCAAAACGAAGAATACTCAACAACAAATGGACAGCAAAATGGATTTAGTCGAATT 2307  
|||||

Db 246 ProIleAsnGluGluIleThrGluGluAsnGlyThrAlaAsnGlyPheSerGluIle 265  
|||||

QY 2308 AACTCAAAAGAGAAAGCTTTGAAAACAGATAGTGTCTCAGGGGTTTCAAAAACAGACGAAA 2367  
|||||

Db 266 AsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSerLysGlnSerLys 285  
|||||

QY 2368 AAACAAAACCTC 2379  
|||||

Db 286 LysGlnLysLeu 289  
|||||

RESULT 5  
US-09-764-864-1369  
; Sequence 1369, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1369  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (9)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (16)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (18)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (31)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (46)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (121)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (283)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-864-1369

Alignment Scores:  
Pred. No.: 6, 05e-118 Length: 292  
Score: 1453.50 Matches: 280  
Percent Similarity: 96.25% Conservative: 2  
Best Local Similarity: 95.56% Mismatches: 9  
Query Match: 18.90% Indels: 2  
DB: 9 Gaps: 1

US-09-874-162A-7 (1-4409) x US-09-764-864-1369 (1-292)

QY 748 AGAAAGGATGTAAAGTTGTCATTAAGCAAGTCCACAGGTAAGAGCAGGTGCCTTTG 807  
|||||

Db 1 LysLysAspValSerCysProIle\*\*\*GlnValProThrGlyLys\*\*\*Gln\*\*\*ProLeu 20

QY 808 ATTCTGACCTCAATCAACAAAACCCGGAATTTCCCGTCCCTTCAGTTTCCAGTAAT 867  
|||||

Db 21 AsnProAspLeuAsnGluThrLysProGly\*\*\*PheProSerLeuAlaValSerSerAsn 40

```
QY 868 GAATTGAACTAGTAACAGCCATATGCTGAAGTCTTACTCGTTGCTATTATAGAGTCACT 927
Db 41 GLUPheGluProSer***SerHisMetVallySerTySerLeuLeuPheArgValThr 60
QY 928 COTCCAGAAAGAGAGATTAAATGAATGATTAATGAGAAACCAATGAATAATATTGAT 987
Db 61 ArgProGlyArgGluPheAsnGlyMetIleAsnGlyGluThrAsnGluAsnIleAsp 80
QY 988 GTCATATGAGAGCTCCAGCCAGAGAAACGAATCGTAGGATGGGAAACACATTT 1047
Db 81 ValAsnGluGluLeuProAlaArgArgLysArgAsnArgGluAspGlyGluYsrPhe 100
QY 1048 GTTGACCAAAATGACAGTATTTGATAAAACAGGGCGCTTACAGCTTTTAGATGGGAAATAT 1107
Db 101 ValAlaGlnMetThrValPheAspLysAsnArgLysGlnLeuLeuAspGlyGluYsr 120
QY 1108 GAAGTAGCCATCCAGAAATGGAGAAATGTCATTAAGCAAGAAAGACACATGGAG 1167
Db 121 ***ValAlaMetGlnGluMetGluGluCysProIleSerLysLysArgAlaThrTIPGlu 140
QY 1168 ACTATTCTTGAGGAAAGAGGCTGCTCCATTCGAACATTTCTCAGGGACCTACGTTG 1227
Db 141 ThrIleLeuAspGlyLysArgLeuProPheGluThrPheSerGlnGlyProThrLeu 160
QY 1228 CAGTTCACTCTTCGTTGACAGGAGAGACCAATGATAAATCTAGCGCTCCTATTGCCAAA 1287
Db 161 GlnPheThrLeuArgTIPThrGlyGluThrAsnAspLysSerThrAlaProIleAlaLys 180
QY 1288 CTTCTTGACCTAGAAATTCAGAGAGTCTCCATCAGGAAACACAGCTGGTTTCAGTTAAA 1347
Db 181 ProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysProGlySerValys 200
QY 1348 CTTACTCAAACTATTGCTGTTAAAGAATCATTGACTCAGATCTACAAACAAAGAAAGAA 1407
Db 201 ProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGlnThrArgLysGlu 220
QY 1408 AAGGATCTCCAAATGAAACCGACAAATTAAGATATTTATCAGTTCTCTATAAC 1467
Db 221 LysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyGlnPheLeuTyAsn 240
QY 1468 AACATACAAAGCAACAACTGAAGCAAGAGATGACCTGCAATTCGCCCTTGGTGTACTCTG 1527
Db 241 AsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysProTIPCysThrLeu 260
QY 1528 AACTCCGCGAACTTATAGTTTACTCAAGCATCTTAAACCTGCGCATAGCAGATTTATC 1587
Db 261 AsnCysArgLysLeuTySerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle 280
QY 1588 TTCAACTATGTTTATC---ATCCAAAAGGTGCTAGGA 1621
Db 281 PheAsn***Val-LeuSerSerGlnLysValLeuGly 292
```

## RESULT 6

```
US-09-764-864-1355
; Sequence 1355, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1355
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
```

```
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (166)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-864-1355
```

```
Alignment Scores:
Pred. No.: 4,2e-73 Length: 175
Score: 940.00 Matches: 172
Percent Similarity: 97.73% Conservative: 0
Best Local Similarity: 97.73% Mismatches: 3
Query Match: 12.22% Indels: 1
DB: 9 Gaps: 0
```

US-09-874-162A-7 (1-4409) x US-09-764-864-1355 (1-175)

```
QY 1447 TTTATCACTTCTCTATTAACAACTAACAAGCAACAACTGAAGCAAGAGATGACCTG 1506
Db 1 PheTyGln***LeuTyAsnAsnThrArgGlnGlnThrGluAlaArgAspLeu 20
QY 1507 CATTGCCCTTGTGTACTCTGAACCTGCCCAACTTTATAGTTTACTCAAGCATCTTAAA 1566
Db 21 HisCysProTIPCysThrLeuAsnCysArgLysLeuTySerLeuLeuLysHisLeuLys 40
QY 1567 CTTGCCATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGAT 1626
Db 41 LeuCysHisSerArgPheIlePheAsnTyValTyHisProLysGlyAlaArgIleAsp 60
QY 1627 GTTCTATCAATGAGTGTATGATCGCTCCATGAGGAATCTCAGGAATCTCAGATATTCATCG 1686
Db 61 ValSerIleAsnGluCysTyArgGlySerTyAlaGlyAsnProGlnAspIleHisArg 80
QY 1687 CAACCTGGATTTGCTTTTAGTCGCAACCGACCATTAAGAGAACACCTATCACACATATT 1746
Db 81 GlnProGlyPheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIle 100
QY 1747 CTTGTGTGAGGCCAAACAGAAACAAACAGCAACATGCTGAATTTCTGAATCTCAAGAT 1806
Db 101 LeuValCysArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAsp 120
QY 1807 GGGGAAGTAGAAGCAAGAAACATATAGTAGTGCCCAATCGTCTGTATTTCATAGT 1866
Db 121 GlyGluValGluGlnGlnArgThrTySerSerGlyHisAsnArg***TyPheHisSer 140
QY 1867 GATACCTGTTACTCTCGTCCCAAGAAATGGAAGTAGATAGTAGAAGCAAGAAAGAT 1926
Db 141 AspThrCysLeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAsp 160
QY 1927 CTTGAATGCTAAGAGAAACCAATTCACAAATTTGAAGAGTTT 1972
Db 161 ProGluTIPLeuArg-***LysProLeuHisLysLeuLysSerPhe 175
```

## RESULT 7

```
US-09-764-864-1334
; Sequence 1334, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1334
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
```

```

; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1334

```

Alignment Scores:

Pred. No.:	6,370-55	Length:	278
Score:	733.00	Matches:	142
Percent Similarity:	98.62%	Conservative:	1
Best Local Similarity:	97.93%	Mismatches:	2
Query Match:	9.53%	Indels:	0
DB:	9	Gaps:	0

US-09-874-162A-7 (1-4409) x US-09-764-864-1334 (1-278)

QY	1	CCCGCCCGCGCTCGAGAGCCGACACAGGGGGGCTCTCGATGTAGCACCATGACACAGCG	60
Db	19	ProAlaArgArgSerGlnAsnArgHisGlnGlyGlySerGlyCysSerThrMetThrGly	38
QY	61	ATCGCGCGCGCTCTCTTCCTCCAAATACCTCCCGATTCCGGGGCTCGGACCTCCACTTC	120
Db	39	IleAlaAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe	58
QY	121	CCACCTCGGCGCATCTCATCGAGCACATCGAGGACCAACCACATCGATACAGATCCACGG	180
Db	59	ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThrAspProArg	78
QY	181	GTTTATGAAAAACAGAAATTACAGCGCCACCTATTGTCCTGAGTTACATAATAGA	240
Db	79	ValLeuGluGlyGlnGluLeuGlnGlnProThrTyrValAlaLeuSerTyrIleAsnArg	98
QY	241	TTTCATGACAGATGTCCTCCCGCGGAGAGCAGAGTCCCTAAAGAAAGAGATTTCAGCCGGAAG	300
Db	99	PheMetThrAspAlaAlaArgGluGlnGluSerLeuIleGlyGlyGlnProIle	118
QY	301	CNCTCGCTGACTCTCCAGCTCAGTCTCTCGAGGGAATGTGTCCACTCCCCACGCCAC	360
Db	119	LeuSerLeuThrLeuSerSerSerValSerArgGlyAsnValSerThrProProArgHis	138
QY	361	AGCAGTGAAGCGCTTACTCCCCCGTGCACCCACCCCATCCCGCTCTCTTCATTCCGC	420
Db	139	SerSerGlySerLeuThr***ProValThrProPheThrProSerSerPheArg	158
QY	421	AGCAGCACTCCGACA	435
Db	159	SerSerThrProThr	163

## RESULT 8

```

US-09-874-162A-2
; Sequence 2, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION:
; APPLICANT: Koontz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JUJAZ1 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-2

```

Alignment Scores:	
Pred. No.:	1.21e-48
Score:	661.00
Percent Similarity:	100.00%
Length:	243
Matches:	128
Conservative:	0

```

Best Local Similarity: 100.00%      Mismatches: 0
Query Match:          8.60%         Indels: 0
DB:                   9             Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-874-162A-2 (1-243)

```

US-09-874-162A-7 (1-4409) x US-09-874-162A-2 (1-243)

Qy	52	ATGACAGGATCGCCGCGCTCTCTTCTTCTCCAAATCCTCGGATTCGGGGCTCGGA	111
Db	1	MetThrGlyIleAlaAlaSerPhePheSerAenThrCysargPheGlyGlyCysGly	20
Qy	112	CTCCACTCCCCCACCTCGCGACCTCATCGAGCACATCGAGACAACACATCGATACA	171
Db	21	LeuHisPheProThrLeuAlaAspleuIleGluHisIleGluAspAenHisIleAspThr	40
Qy	172	GATCCACGGGTTTTAGAAAAACAAGAATTACAGCAGCAACCTATGTTGCCCTGAGTTAC	231
Db	41	AspProArgValIeuGluLysGlnGluIeuGlnGlnProThrTyrrValAlaIeuSerTyrr	60
Qy	232	ATAAATAGATTCATCAGACAGATGCTGCCGCGAGAGAGAGTCCCTAAGAAGAAGATT	291
Db	61	IleAsnAArgPheMetThrAspAlaAArgGluGlnGluSerLeuLysIysLysIle	80
Qy	292	CAGCCGAGACTCTCGCTCACTCTGTCAGACTCAGTGTCTCGAGGGAAATGTGCCACTCCC	351
Db	81	GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro	100
Qy	352	CCACGCCACAGCAGTGGAGCTTTACTCCCGCGTGACCCGCCACCCATCACCCCTCCTCT	411
Db	101	ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSerSer	120
Qy	412	TCATTCCGCGAGGACCTCCGACA	435
Db	121	SerPheArgSerSerThrProThr	128

## RESULT 9

```

US-10-144-198-16
; Sequence 16, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-16

```

**Alignment Scores:**

Pred. No.:	1,218-48	Length:	243
Score:	661.00%	Matches:	128
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.60%	Indels:	0
DB:	15	Gaps:	0

US-09-874-162A-7 (1-4409) x US-10-144-198-16 (1-243)

52	ATGACAGCATCGCGCGCGCTCTCTTCTTCCAAATACCTCGATTCCGGGGCTCGGA	111	QY
1	MetThrGlyIleAlaAlaAlaSerPhePheSerAsnThrCysargPheGlyGlyCysGly	20	Db
112	CTCCACTCCCAACCTGGCGGACCTCATCGAGCAATCGAGACCAACCATCGATACA	171	QY
21	LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr	40	Db
172	GATCCACGGGTTTAGAAAAACAAGATTACACAGCCCAACCTATGTTGCCCTGAGTTAC	231	QY
41	AspProargValLeuGluLysGlnLeuGlnGlnProThrTrpValAlaLeuSerTrp	60	Db



QY 232 ATAAATAGATTTCATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAAGAAAGAGATT 291  
 Db 61 IleasnArgPheMetThrAspAlaAlaArgGlnGlnGluSerLeuLysLysLysLe 80  
 QY 292 CAGCCGAAGCTCTCGCTGACTCTCTCCAGCTCAGTGTCTCGAGGAATGTCCACTCCC 351  
 Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100  
 QY 352 CCAGCCACACAGTGGNAGCTTACTCCCGCCGAGAGCAGGAGTCCCTAAAGAAAGAGATT 411  
 Db 101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSerSer 120  
 QY 412 TCATTCCGCGACGACTCCCGACA 435  
 Db 121 SerPheArgSerSerThrProThr 128

RESULT 10  
 US-10-104-047-2404  
 ; Sequence 2404, Application US/10104047  
 ; Publication No. US20030236392A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA  
 ; FILE REFERENCE: H1-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047  
 ; CURRENT FILING DATE: 2002-03-25  
 ; PRIOR APPLICATION NUMBER:  
 ; PRIOR FILING DATE:  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 2404  
 ; LENGTH: 243  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-104-047-2404

Alignment Scores:  
 Pred. No.: 1,21e-48 Length: 243  
 Score: 661.00 Matches: 128  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8.60% Indels: 0  
 DB: 15 Gaps: 0

US-09-874-162A-7 (1-4409) x US-10-104-047-2404 (1-243)

QY 52 ATGACAGGATCGCCGCGCTCTCTTCTCCATACCTCCGAGTCCGGGGTCCGGA 111  
 Db 1 MetThrGlyIleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly 20  
 QY 112 CTCACCTTCCACCTCGCGACCTCATCGAGCAGCATCGAGGACCAACCATCGATACA 171  
 Db 21 LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr 40  
 QY 172 GATCCACGGGTTTATAGAAAAACAAGATTACAGAGCCAACTATGTCCTCCGAGTTAC 231  
 Db 41 AspProArgValLeuGluLysGlnGlnGlnProThrTyrValAlaLeuSerTyr 60  
 QY 232 ATAAATAGATTTCATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAAGAAAGAGATT 291  
 Db 61 IleAsnArgPheMetThrAspAlaAlaArgGlnGlnGluSerLeuLysLysLysLe 80  
 QY 292 CAGCCGAAGCTCTCGCTGACTCTCTCCAGCTCAGTGTCTCGAGGAATGTCCACTCCC 351  
 Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100  
 QY 352 CCAGCCACACAGTGGNAGCTTACTCCCGCCGAGAGCAGGAGTCCCTAAAGAAAGAGATT 411  
 Db 101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSerSer 120  
 QY 412 TCATTCCGCGACGACTCCCGACA 435

Db 121 SerPheArgSerSerThrProThr 128

RESULT 11  
 US-09-764-864-884  
 ; Sequence 884, Application US/09764864  
 ; Patent No. US20020132753A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT223  
 ; CURRENT APPLICATION NUMBER: US/09/764,864  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1792  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 884  
 ; LENGTH: 143  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-764-864-884

Alignment Scores:  
 Pred. No.: 3,05e-40 Length: 143  
 Score: 564.00 Matches: 109  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.09% Mismatches: 0  
 Query Match: 7.33% Indels: 0  
 DB: 9 Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-884 (1-143)

QY 1 CCGCGCCGCGCTCCGAGCCGACACAGAGGGGCTCTCGATGTAGCACCATGACAGCC 60  
 Db 34 ProAlaArgArgSerGlnAsnArgHisGlnGlyGlySerArgCysSerThrMetThrGly 53  
 QY 61 ATCGCGCGCGCTCTCTTCTTCCAAATACCTGCGGAGTCCGGGGTCCGAGTCCACTTC 120  
 Db 54 IleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe 73  
 QY 121 CCCACCTGCGGAGCTCATCGAGCAGCATCGAGGACCAACCATGATACAGATCCACGG 180  
 Db 74 ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThrAspProArg 93  
 QY 181 GTTTTAGAAAAACAAGATTACAGAGCCAACTATGTCCTCCGAGTTACATAAATAGA 240  
 Db 94 ValLeuGluLysGlnGlnGlnProThrTyrValAlaLeuSerTyrIleAsnArg 113  
 QY 241 TTCATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAAGAAAGAGATTCCGCGAG 300  
 Db 114 PheMetThrAspAlaAlaArgGlnGlnGlnGluSerLeuLysLysLysLysGlnProLys 133  
 QY 301 CTCCTGCTGACTCTGTCCTCCAGCTCAGTGTCT 330  
 Db 134 LeuSerLeuThrLeuSerSerSerValSer 143

RESULT 12  
 US-10-029-386-28495  
 ; Sequence 28495, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEOMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 28495  
 ; LENGTH: 69  
 ; TYPE: PRT



```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P23477, EVALUE 2.70e+00
US-10-029-386-28495

Alignment Scores:
Pred. No.: 2,01e-22 Length: 69
Score: 359.00 Matches: 68
Percent Similarity: 98.55% Conservative: 0
Best Local Similarity: 98.55% Mismatches: 1
Query Match: 4.67% Indels: 0
DB: 14 Gaps: 0

US-09-874-162A-7 (1-4409) x US-10-029-386-28495 (1-69)
QY 1756 AGGCCAAAGCAACAAAGCAAGCATCTGTAATTTCTGTAATCTGAAGTGGGAAGTA 1815
Db 1 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 20
QY 1816 GAACAGCAAGAACATATAGTAGTGCCCAATCGTCTGTATTTCCATAGTCATCTGC 1875
Db 21 GluGlnArgThr-Tyr-Ser-SerGlyHisAsnArgLeuTyr-PheHisSerAspThrCys 40
QY 1876 TTACTCTCCGTCACAGAAATGGAAGTAGTAGTAGAGAGTGAAGTGAAGTGAAGTGAAGTGG 1935
Db 41 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 60
QY 1936 CTAAGAGAAAAACCATACACAAATT 1962
Db 61 LeuArgGluLysThrIleThrValIle 69

RESULT 13
US-10-276-774-2060
; Sequence 2060, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tansg, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276, 774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2060
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2060

Alignment Scores:
Pred. No.: 2,27e-15 Length: 171
Score: 280.00 Matches: 56
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 12 Gaps: 0

US-09-874-162A-7 (1-4409) x US-10-276-774-2060 (1-171)
QY 258 CAGGAGTCCCTAAAGAAAGATTGAGCGGAGCTCTCGTACTCTCTCAGTCTCAGT 327
Db 1 GlnGluSerLeuLysLysLysIleGlnProLysLeuSerLeuThrLeuSerSerVal 20

```

```

QY 328 TCTCAGGGAATGTCTCCTCCTCCCTCCCGCAGCCAGCAGTGGAGAGCCTTACTCCCCCGTG 387
Db 21 SerArgGlyAsnValSerThrProArgHisSerSerGlySerLeuThrProProVal 40
QY 388 ACCCCACCCATCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 435
Db 41 ThrProProLysThrProSerSerSerPheArgSerSerThrProThr 56

RESULT 14
US-10-231-778-2
; Sequence 2, Application US/10231778
; Publication No. US20030126647A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau, Pierre
; APPLICANT: Chaudhury, Abdul M.
; APPLICANT: Dennis, Elizabeth S.
; APPLICANT: Koltunow, Anna M.G.
; APPLICANT: Luo, Ming
; APPLICANT: Peacock, William J.
; TITLE OF INVENTION: Method for inducing seed development by down-regulating
; TITLE OF INVENTION: expression of the FIS2 gene
; FILE REFERENCE: 72-98A
; CURRENT APPLICATION NUMBER: US/10/231,778
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: 09/398,237
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 60/101,184
; PRIOR FILING DATE: 1998-09-21
; PRIOR APPLICATION NUMBER: AU PP6061
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU PP6062
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU PP6063
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU P01345
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: AU P01346
; PRIOR FILING DATE: 1999-07-01
; NUMBER OF SEQ ID NOS: 239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-231-778-2

Alignment Scores:
Pred. No.: 2,14e-09 Length: 813
Score: 214.50 Matches: 147
Percent Similarity: 32.70% Conservative: 113
Best Local Similarity: 18.49% Mismatches: 298
Query Match: 2.79% Indels: 237
DB: 14 Gaps: 28

US-09-874-162A-7 (1-4409) x US-10-231-778-2 (1-813)
QY 436 GAGCCACACAGATCTATAGATTTCTTCCGAATCTCGGAATCTCATAGCACCATTATTTTG 495
Db 70 LysProValAlaValTyrLysLysLeuGluThrArgSerLysAsnAsnProTyrPheLeu 89
QY 496 CACAGAACTCTTACTACATG-----TCTCATCGAACTCCAGAACAAACATCAAAAGG 549
Db 90 ArgArgSerLeuLysTyrIleIleGlnAlaLysLysLysLysSerAsnSerGlyGly 109
QY 550 AAA---ACATTAAAGTTGATGATATGTTTATCAAAAGTAGAGAAAATGAAAGGAGAGCAA 606
Db 110 LysIleArgPheAsnTyrArgAspValSerAsnLysMetThrLeuLysAlaGluValVal 129
QY 607 GAATCTCATGAC-----TTGTCAGCT 627
Db 130 GluAsnPheSerCysProPheCysLeuIleProCysGlyGlyHisGluGlyLeuGlnLeu 149

```



```

; Sequence 6, Application US/10177478
; Publication No. US20030165903A1
; GENERAL INFORMATION:
; APPLICANT: Dang, Van-Dinh
; APPLICANT: Okamuro, Jack
; TITLE OF INVENTION: Chimeric Histone Acetyltransferase
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 11696-037001
; CURRENT APPLICATION NUMBER: US/10/177,478
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: U.S. 60/300,135
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 692
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-177-478-6

```

```

Alignment Scores:
Pred. No.: 4,09e-09 Length: 692
Score: 211.00 Matches: 149
Percent Similarity: 32.46% Conservative: 122
Best Local Similarity: 17.84% Mismatches: 286
Query Match: 2.74% Indels: 278
DB: 14 Gaps: 30

```

US-09-874-162a-7 (1-4409) x US-10-177-478-6 (1-692)

```

QY 106 TCGGAGCTCCACTTCCCGACCTCGGACCTCATCGAGCATCGAGGACCAACAC--- 162
DB 20 CysGlyGlyHis-----GluGlyLeuGlnLeuHisLeuLysSerSerHisAsp 35
QY 163 -----ATCGATACAGATCCACGGGTT----- 183
DB 36 AlaPheLysPheGluPheTyrArgAlaGluLysAspHisGlyProGluValAspValSer 55
QY 184 -----TTAGAAAACAGATTACACGCCACCTAT 216
DB 56 ValLysSerAspThrIleLysPheGlyValLeuLysAspValGlyAsnProGlnLeu 75
QY 217 GTTGCCTCGAGTTACATAAATAGATTTCATGACAGATGTCGCCGCCGAGAGCAGGATCC 276
DB 76 SerProLeuThrPheCysSerLysAsnArgAsnGlnArgGlnArgAspSerAsn 95
QY 277 CTAAGAGAGAGATTCAG-----CCGAGGCTC 303
DB 96 AsnValLysLysLeuAsnValLeuLeuMetGluLeuAspLeuAspLeuProArgGly 115
QY 304 TCGCTGACTCTCTCCAGCTCACTGCTCGAGGAGATGTCTCACTCCCGCCAGC---CAC 360
DB 116 ThrGluAsnAspSerThrHisValAsnAspAspAsnValSerSerProArgAlaHis 135
QY 361 AGCACTGGAAGCTTTACTCCCGCTGACCCGCCATCCACCCCTCTCTTCATTCGCG 420
DB 136 SerSerGluLysIleSerAspIleLeuThr-----ThrGlnLeuAlaIleAlaGlu 153
QY 421 AGCAGCATCCGACAGAGCCACAGATCTATAGATTCTTCGAACTCGGAATCTATA 480
DB 154 SerSerGluProLysValPro-----HisValAsnAspGlyAsnValSer 168
QY 481 GCACCAATATTTTGCACAGAACTCTTACTATCATGTCTCATGAACTCCAGCAACAA 540
DB 169 SerProProArgAlaHisSerSer-----AlaGluLysAsnGluSerThrHis 184
QY 541 ATCAAAAGGAAACATTTAAAGTTGATGATATGTTATCAAAAGTAGAAATGAAAGGA 600
DB 185 ValAsn-----AspAsp-----AspValSerSer 193
QY 601 GAGCAAGATCTCATAGCTTGCACCTCATTTGGACGCTTACGTTTACTGGTTTCTTCCAC 660
DB 194 ProProArgAlaHisSerLeu-----Glu 201

```

```

QY 661 AAAAATGATAAGCCATCACCAAACTCAGAAAAT----- 693
DB 202 LysAsnGluSerThrHisValAsnGluAspAsnIleSerSerProProLysAlaHisSer 221
QY 694 ---GAACAAAATTCCTGTACCTCGGAAGTCCTCTGTGTGAAGTTTGCACAAAAGAA 750
DB 222 SerLysLysAsnGluSerThr-----HisMetAsnAsp 232
QY 751 AAGGATGTAAGTTCTCAATAAGG-----CAAGTTCCCA 786
DB 233 GluAspValSerPheProProArgThrArgSerSerLysGluThrSerAspIleLeuThr 252
QY 787 GGTAAAGACGAGGTGCTTCGATTCCTGACCTCAATCAAAACAAACCCGGAAT----- 840
DB 253 ThrThrGlnProAlaIleValGluProSerGluProLysValArgGlySerArgArg 272
QY 841 -----TTCCCGTCCCTTGCAGTTTC 861
DB 273 LysGlnLeuTyrAlaLysArgTyrLysAlaArgGluThrGlnProAlaIleAlaGluSer 292
QY 862 AGT-----AATGAATTTGAACCTAGTAACAGCCAT 891
DB 293 SerGluProLysValLeuHisValAsnAspGluAsnValSerSerProProGluAlaHis 312
QY 892 ATGCTGAAGTCTTACTCGTTGCTATTATTAGAGTCACTCGTCCAGGA----- 936
DB 313 SerLeuGluLysAlaSerAspIleLeuThrThrGlnProAlaIleAlaGluSerSer 332
QY 937 -----AGAAGAGAGTTAAT 951
DB 333 GluProLysValProHisValAsnAspGluAsnValSerSerThrProArgAlaHisSer 352
QY 952 GGAATGATTAAATGAGAAACCAATGAAATATTGATGTCAATGAAGAGCTTCCAGCCAGA 1011
DB 353 SerLysLysAsnLysSerThrArgLysAsnValAsp---AsnValProSerProLys 371
QY 1012 AGAAACGAAATCGTGAGGATGGGAAACACATTTGTCACAAATGACAGTATTGAT 1071
DB 372 ThrArgSerSerLysLysThrSerAspIleLeuThrThrGlnProThrIleAlaGlu 391
QY 1072 AAAACAGGGCTTACAGCTTTTAGTGGGAATATAGTAGCCATGAGCAAGAAATGAA 1131
DB 392 SerSer-----GluProLysValArgHisValAsn 401
QY 1132 GAATGCTCAATAAGCAAGAAAGAGCAACATGGGAGACTATTCTTGATGGGAAGGCTG 1191
DB 402 AspAspAsnValSerSer----- 407
QY 1192 CCTCCATTGCAACATTTTCTCAGGACCTACCTTCAGTTCAGTTCTCGTTCGACAGGA 1251
DB 408 -----ThrProArgAlaHisSer 413
QY 1252 GAGACCAATGATAAATCTACGGCTCTCTATTGCCAAACCTCTTGCCACTAGAAATTCAGAG 1311
DB 414 SerLysLysAsnLysSerThr-----ArgLysAsnAspAsp 425
QY 1312 AGTCTCCATCAGAAACAGCCCTGTTGATTAAACCTACTCAAACTATTGCTGTATAA 1371
DB 426 AsnIleProSerProProLysThrArgSerSerLysLysThrSerAsnIleLeuThrArg 445
QY 1372 GAATCATTCAGTACAGATCTACAAACAGAAAGAAAGGATACCTCAAAATGAAACCGA 1431
DB 446 -----ThrGlnProAlaIleAlaGluSerGluProLysValProHisValAsnAsp 462
QY 1432 CAAAATTAAGAATATTTATAGTTTCTTATACAGATCTATACAAATACAGGCAACAACTGAA 1491
DB 463 AspLysValSerSerThrProArg---AlaHisSerSerLysLysAsnLysSerThrHis 481
QY 1492 GCAGAGATGAC---CTGCATTGCCCTTGGTGACTCTGAACCTGCGCGCAACCTTATAGT 1548
DB 482 LysLysAspAsnAlaSerLeuProLysThrArgSerSerLysLysThrSerAsp 501

```

```
QY 1549 TTACTCAAGCATCTTAAACTCTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCA 1608
Db      :|||
502 IleLeu----- 503
QY 1609 AAAGGTCTAGGATAGATGTTTCTATCAATGAGTGTATGATGCTCTATGCAGAAAT 1668
Db      :|||
503 ----- 503
QY 1669 CCTCAGGATATTCATCGCAACCTGGATTGCTTTAGTCGCAACGGACCGATTAAAGAGA 1728
Db      :|||
504 -----AlaThrThrGlnProAlaLysAla 511
QY 1729 ACACCTATCACACATATTCTTGTGTGTCAGGCGAAACGACAAAGCAAGCATGCTCTGAA 1788
Db      :|||
512 GluPro-----SerGluProLysValThrArgValSerArgArgLys 525
QY 1789 TTCTTGAATCTGAAGATGGGAAGTAGACACAGCAAGAACATATAGTAGTGGCCACAAT 1848
Db      :|||
526 GluLeuHisAlaGluArgCysGluAlaLysArgLeuGluArgLeuLys-----Gly 542
QY 1849 CGTCTGTATTCCATAGTAGTACCTGCTTACCTCTCGTCCACAGAAATG-----GAA 1902
Db      :|||
543 ArgGlnPheTyrHisSerGlnThrMetGlnProMetThrPheGluGlnValMetSerAsn 562
QY 1903 GTAGATAGTGAAGATCAAAAGATCCTGATGGCTAGAGAAAAACCATTTACACAAAT 1962
Db      :|||
563 GluAspSerGluAsnGluThrAspAspTyrAlaLeuAspLysSerGluArgLeuArgLeu 582
QY 1963 GAAGAGTTTCTGATGTTTAATGAGGAGAGAGAAAGTAGTGAACCTCGAATCTCCAT 2022
Db      :|||
583 GluArgLeuValGlyValSerLysGluLysArgTyrMetTyrLeuTrpAsnIlePhe 602
QY 2023 GTCATGAAGCATGGGTTTATTGCTGACAAATCAATGAATCATGCTGTATGCTGTTTGA 2082
Db      :|||
603 ValArgLysGlnArgValIleAlaAspGlyHisValProTrpAlaCysGluGluPheAla 622
QY 2083 GAA-----AATTATGGACAGAAAAATTAAT 2106
Db      :|||
623 LysLeuHisLysGluGluMetLysAsnSerSerPheAspTrpTrpArgMetPhe 642
QY 2107 AAGAGAATTTATGCGAACTTCATGCTTCTATGTCAGCATGCGATGACTTTAATCTT 2166
Db      :|||
643 ArgIleLysLeuTrpAsnAsnGlyLeuIleCysAlaLysThrPheHisLysCysThrThr 662
QY 2167 ATTACATAATGTCATATAGATAAAGCTGTACCAAGCTCCGTAAGATGCAGCAAAATTA 2226
Db      :|||
663 IleLeuLeuSerAsnSerAspGluAla-----GlyGlnPhe 674
QY 2227 GAAAGGGGGAATCTGCTTCCCTGCAACGAGAAATACTGAA 2271
Db      :|||
675 ThrSerGlySerAlaAlaAsnAlaAsnAsnGlnGlnSerMetGlu 689
```

Search completed: August 25, 2004, 20:11:22  
Job time : 479.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 25, 2004, 18:23:29 ; Search time 138.5 seconds  
(without alignments)  
6124.310 Million cell updates/sec

Title: US-09-874-162A-7  
Perfect score: 7690  
Sequence: 1 cccgcgcgcgcgcgcgcag.....tacataaagttcaataatat 4409

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cn2 1/USPTO spool\_p/US09874162/runat\_25082004\_171844\_2190/app query.fasta\_1.4551  
-DB=PIR\_78 -QFMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=BITS -START=1 -END=1 -MATRIX=blusun62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874162 @CN\_1\_1\_218 @runat\_25082004\_171844\_2190 -NCPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	188	2.4	2013	2 C71610	probable membrane
2	178.5	2.3	1314	2 T09481	mating type silenc
3	176.5	2.3	1142	2 S59359	GIN4 protein - yea
4	166	2.2	1507	2 T42631	breast cancer tumo
5	161	2.1	1875	2 S38173	myosin-like protei
6	159.5	2.1	912	2 B44450	ubiquitin-specific
7	159.5	2.1	1263	2 T15496	hypothetical prote
8	158	2.1	1119	2 B70126	surface-located me
9	157	2.0	1208	2 T39068	coiled coil protei
10	156	2.0	1755	2 S69388	TYB protein - yea
11	155.5	2.0	1755	2 S57045	TYB protein - yea
12	155.5	2.0	3724	2 T18427	hypothetical prote
13	154	2.0	1271	2 A45555	glutamate rich pro
14	153	2.0	707	2 A46302	PTB-associated spl

15	152	2.0	1749	2 S69972	TYB protein - yea
16	152	2.0	1755	2 S57047	TYB protein - yea
17	152	2.0	1937	2 T38077	hypothetical coile
18	151	2.0	1328	2 S52894	TYB protein - yea
19	151	2.0	1755	2 S69969	TYB protein - yea
20	150.5	2.0	1368	2 T18371	probable glutamate
21	150.5	2.0	1668	2 T13748	sex comb proline-r
22	149	1.9	891	2 G84693	probable proline-r
23	149	1.9	1145	2 T18235	transcription acti
24	149	1.9	1295	2 T24587	hypothetical prote
25	149	1.9	1737	2 A59235	unconventional myo
26	148	1.9	863	2 T51002	hypothetical prote
27	148	1.9	1328	2 B28097	TYB protein - yea
28	148	1.9	1604	2 S70227	TYB protein - yea
29	148	1.9	1755	2 S69957	TYB protein - yea
30	147.5	1.9	1755	2 S69980	TYB protein - yea
31	147.5	1.9	1730	2 S67593	transport protein
32	147	1.9	1085	2 S62516	hypothetical coile
33	147	1.9	1196	2 S40908	TYB protein - yea
34	147	1.9	1328	2 S52481	TYB protein - yea
35	147	1.9	1755	2 S50641	TYB protein YER138
36	147	1.9	1755	2 S69975	TYB protein - yea
37	146.5	1.9	1235	1 PMBYH	potassium transpor
38	146	1.9	1731	2 T20815	hypothetical prote
39	145	1.9	1744	2 JH0720	tanabin - African
40	145	1.9	1755	2 S50563	TYB protein - yea
41	145	1.9	1953	2 S63244	BNIL protein - yea
42	145	1.9	3328	2 T30835	breast cancer tumo
43	144.5	1.9	1738	2 T14867	interaptin - slime
44	144	1.9	1613	2 D90129	hypothetical prote
45	144	1.9	1755	2 S45736	TYB protein - yea

ALIGNMENTS

RESULT 1  
C71610  
Probable membrane associated protein PFB0615c - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C/Accession: C71610  
R: Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
R: Berteaux, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;  
Science 282, 1126-1132, 1998  
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A/Reference number: A71600; MUID:99021743; PMID:9804551  
A/Accession: C71610  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-2013 <GAR>  
A/Cross-references: GB:AE001406; GB:AE001362; NID:G3845230; PIDN:AACT1912.1; PID:G384523  
A/Experimental source: clone 3D7  
C/Genetics:  
A/Gene: PFB0615C

Alignment Scores:			
Pred. No.:	1.08e-05	Length:	2013
Score:	188.00	Matches:	236
Percent Similarity:	33.41%	Conservative:	183
Best Local Similarity:	18.82%	Mismatches:	362
Query Match:	2.44%	Indels:	473
DB:	2	Gaps:	60

US-09-874-162A-7 (1-4409) x C71610 (1-2013)	
Qy 688 GAAATGACAAATTCCTGTTGAGAGTCTGTTGAGAGTTCCACAAAAA 747	
Db 69 GluAspCysAsnLysAsnMetThrIleAsnMetLeuGlnAsnGluAsnLys 88	
Qy 748 AGAAGATGATAGTTGTCCTCAATAGGCAAGTCCACAGGTAAAGACGAGTTC 807	
Db 89 IleIleLysGluThrCysTyrIleLysAsnValThrAsnLysIleTyrHisThrLeu 108	



684 -----AsnLeuHisLysTyr----- 688

2650 TGTCATCAAAATATGTTGAATGTACTAGAATATATTTTCATATATAAATCAGAATTTTTTT 2709

689 -----IleLeuAsnAsnIleThrSerPheTyrIleMetSerLysTyrLeuPheV 705

2710 GCATTTATGAACGCTGTTTTCTACCTTTGTAATGTGAGACATTTTCTTGGGGAGGAA 2769

705 alLeuLeuAspAsnGlyLeuLeuTyrTyrThrLysLysAsn-----AspAspGlyL 722

2770 AATTGGAAATGGTTCCCTTTTGAAGATTCGAAGTGGCTTCATATGTCCAACTACAGAAAA 2829

722 ySile-----TyrAspPheLeuGluLeuSerAsnPheTyrIleCysTyr-TyrLysAsn 739

2830 GGAAAAAATAGAAATTCAGAGATTTTTTGAATATATATGCAAT-----TACTATTG 2883

740 IleAsn-----LysIleValAspIleLysIleIleAsnGluHisAspIleTyrTyrMet 757

2884 CAGTCAAACTTGATCCTTGTTTTGAAATCATTTGTCAATTCCGAATGAAAAATATATA 2943

758 AspLys-----LysHisIle-----LeuLysAsnHisSerLeuLys 769

2944 TGTAATTTTACATTCATATAAGTTCTCTTTTACAATATAAAAAATAGCACTTCTTCATCTTAT 3003

770 AsnAsnTyr-----LeuAsnIleIleAsnThrLysGluLys----- 781

3004 GCCTGTTTGAGAAGATATTAATTTTTCACATCTGTG--ACAGTGAATGCTATGTGGT 3060

782 -----IleGlnSerTyrAsnIlePheSerMetLeuGluAsnCysThrCys----- 796

3061 TTATAAGATTACAGACCATTTGTTTTTCATGTGGATATTTTAGTGCATTGCTCACCGGT 3120

797 -----IlePheLeuSerLeuAsnAspGly 804

3121 ATGTTTTTTTTTTTAACTTTGAACATTTTGGCTTTGTTTTGTTTTCTTTTAAATAGATA 3180

805 SerPheTyrPheIleAsn----- 811

3181 ATCACACGGAAAATTAAGCTGTTTCATATCTTTAAATTAGGATTGCAACCAAGGAAGAA 3240

812 ThrLysHisLysIleLeuLeuTyrGluAsnLeuGln-AsnPheSerAsnLeuGlyHisAs 831

3241 CGCATTTGAGATTTTAAGATGTCACTTATAAGGGGAGAAGTGTCTTAAAGATCAACCA 3300

831 nGlnIle-----TyrCysAsnPheLysLysAsnLys 841

3301 GAAAACTGTTATGCCCTTTTATTTGTTGCAAGGATGCTTTGTAATGTGTTTCATCAATA 3360

841 s-----TyrIleGlnTyrSerValPheAsnLysLeuAsnGlu-- 853

3361 GAATATCCAATAGAGATAAGCTGACTTGAATCAATTTTGTAGCAATTTTGCCCTGTGTATA 3420

854 -----Tyr 854

3421 TGTGTTTTCGCACATATTTGCAGTTGGATTTTCTCAACAGAAAGTGGATTCACACTG 3480

854 rIlePheAsnGlyTyr-----PheTyrValGlnGlnTyrIleIlePhePhe 869

3481 GCACATT-----AACAAAGCACCAATAGGTTTTTTATTCCAACTCCGAGCACTGTGT 3531

869 ePheLeuIleTyr-SerThrSerHisLysLysPhePheIle-----TyrLeuVa 885

3532 TGAGTACATCACTCACTCAATTTTTTATATCTTTAAGATATTCATTTTCATATCTTTTA 3591

885 lGluAsnIleHisIle-----TyrIleLeuPheLysLysIleHisGlnThrAsnIleLeuTy 904

3592 TTTTATAAGG--ATCAATGCTGCTGTAATAACAGGTATTTTAAATTTTAAAAATTCATT 3648

904 rLysAsnLysGluLysAsnGlnAsnGlnAsnGluAsnIleIleAsnMetLys----- 921

3649 CCACCCATCAGATGCGAGTCCCTATTTTGTTTTATGAAGGATATATAAGCTTTCTATA 3708

922 -ArgGlnLysGluSerSerAsnTyrIleLeuTyrAsnPheTyrAsnPheTyrLys----- 938

[illegible]

QY 202 ---CAGCAGCAACCTATTGTCCTGAGTTTACATAAATAGATTATCATGACAGATGCTGCC 258  
 Db 190 ArgGlnGlnSerSerLeuSerSerGlyArgIleHisLysHisIleGluSerProPhe 209  
 QY 259 CCCCAGAGCAGGAGTCCCTAAAGAG-----AAGATTACGCGGAAGCTTCGCTG 309  
 Db 210 GlnArgAsnSerLysProValSerLysSerIleValLysAlaProAsnLysValAsnPhe 229  
 QY 310 ACTCTGTCCAGTCACTGCTC-----CGAGGGAATGTCTCCCTCCCGACGACG 363  
 Db 230 AlaPheGluThrSerLeuAsnLysPheAspLeuLeuThrArgAlaArgAlaSer 249  
 QY 364 AGTGAAGCCTTACTCCCGCGTGACCCACCCATCCCTCTCTCTTCAATCCGCGAGC 423  
 Db 250 SerGlyGly-----ThrProThrSerLysGlySerAla 260  
 QY 424 AGCACTCCGACAGAGCCACACAGATCTATAGATTCTTCAACTCGGAAT----- 474  
 Db 261 AlaValProAlaAlaPro-----LysThrLysAsnLysThrAsp 273  
 QY 475 -----CTCATAGCACCAATATTTTTCACAGAACTCTTACTTACATGCTCAT 522  
 Db 274 SerGlnValSerLysIleAlaAsnGluIleGlnAsnSerAsnLysSerArgIleSerLys 293  
 QY 523 CGAACTCCGACAGCAACATCAAAAGGAAACATTTAAAGTTGATGATATGTTATCAAAA 582  
 Db 294 AlaAsnSer-----ProIleGlnArgThrThrThrThrSerAsnIleProSerGlu 311  
 QY 583 GTAGAGAAATGAAAGGAGAGCAAGATCTCATAGCTTGTGAGCTCATTTGACGCTTACG 642  
 Db 312 AlaLysGluLeu-----TyrAsnMetLeuSerSerArgLeuLysArgArg 326  
 QY 643 TTTACTGTTTCTCCACAAAATGATAAGCCATCACCAAACTCAGAAATGAACAAAAT 702  
 Db 327 AsnValLeuPheAspLysLysGlyThrIleAsnSerGluAsnAsnAspLysThrGlyAla 346  
 QY 703 TCTGTTACCTCGAGTCTGCTGTGAAAGTTTCCACAAAAAGAAAGAGATGAAGT 762  
 Db 347 SerValLysValGluSerSerHisValGluIleIleGlySerLysArgAspHisThrHis 366  
 QY 763 TGTCCTAATAAGCAAGTCCACACAGTAAAGCAGGTGCTTTGATTCTGACCTCAAT 822  
 Db 367 ThrSerPheMetGluLeuProSerMetGlnAlaThrIleAlaAspLeuAlaLysThrThr 386  
 QY 823 CMAACAAACCCGGAAATTCCTGCTCCCTGAGTTTCCAGTATGATGATTTGAACCTAGT 882  
 Db 387 SerLeuValLysGluAsnSerAlaGlnIleGluValAspArgAsnLeuIleGlyProGlu 406  
 QY 883 AACACCATATGTTGAAGTCTTACTCGTTGCTATTTAGATGACTCTGTCAGGAGGAAGA 942  
 Db 407 ArgSer-----ArgLysAsnGluLys 413  
 QY 943 GAGTTTAATGAATGATTATGAGAAACCAATGAAATATGATGTCATGAAGACGCTT 1002  
 Db 414 GluPheGlnAspLeu-----GluThrArgThrSerSerProAlaLysGluSerLeu 430  
 QY 1003 -----CCAGCCAGAGAAACCGAAATCTGAGGATGGGGAAGAAG 1041  
 Db 431 AlaSerProSerAlaAspThrProGluArgLysValProGlnLeuAspLeuThrSerPro 450  
 QY 1042 ACATTGTTGCAAAATGACAGTATTTGATATAAAACAGCGGCTTTACAGCTTTTATGATGGG 1101  
 Db 451 SerPheAspSerThrProGlyValThrGluLysMetAspLeuLysGluValProSerPro 470  
 QY 1102 GAATATGAA-----GTAGCCATGAGGAATGGGAAGATGTCCTCAATAGC--- 1146  
 Db 471 GlutryrProSerSerAlaGlyClnAlaValGlnGlnSerGluSerSerGlnIleSerHis 490  
 QY 1147 -----AAG 1149  
 Db 491 ValIleGlnGlnAlaGluAlaAspSerGlySerHisGlyIleGlnValGluSer 510

QY 1150 AAAAGAGCAACATGGAGAGCTATTCTTGATGGAGAGAGGCTGCCCTCCATTCGAAACATTT 1209  
 Db 511 SerGluLeuAsnIleProThrIlePheGluLysLysAspAsnPro----- 525  
 QY 1210 TCTCAGGACCTACGTTGCGATTCCTCTTCCTGGACAGGAGACCAATGATAAATCT 1269  
 Db 526 ---GlnAlaAlaSerAlaSerGlyThrPheHisHisValAspGlu-----LysGlnSer 542  
 QY 1270 AGCGTCTCTATTGCCAAACCTCTGCGCACTAGAAATTCAGAGAGTCTCCATCAGGAAAC 1329  
 Db 543 ThrGluGluThrAlaLysLeuAlaLeuSerAspThrSerSer-----GluThrSer 560  
 QY 1330 AGCCTGTTTCA-----GTTAAACCTACTCAAACTATTGCTGTTTAAAGNA 1374  
 Db 561 SerAspGlySerSerSerGlyHisTyrValGlyAspPheGlnLeuLeuAsnThrMetLys 580  
 QY 1375 TCATTGACTACAGATCTACAAACAAAGAAAGAAAGGATCTCCAAATGAAACCGACAA 1434  
 Db 581 AlaArgMetValGlyProLysValAlaLysTyrLysAlaPhe-----AsnLysSer 597  
 QY 1435 AAATTAAAGAAATTTTATCAGTTTCTCTATAACAACAATACAAAGCAACAACACTGAAGCA 1494  
 Db 598 LysPheLysAspLeuMetLysTyrLeuIleAlaGluHisThr-----ProLeuAlaAla 615  
 QY 1495 AGAGATGACTGCTCATTGCGCTTGGTGTACTCTGAACCTGCGCAAACTTTATAGTTTA--- 1551  
 Db 616 ArgSerAspValSerCysSerAsnAspIleCysAsnIleArgAsnTyrGluSerValGlu 635  
 QY 1552 CTCAGCACTTTAAACTCTGCCATAGCAGATTTATCTTC----- 1590  
 Db 636 ValHisHisLeuGlnLysLeuProArgLeuArgTyrLeuLeuLysLeuSerProSer 655  
 QY 1591 ---AACTATGTTATCATCCAAAAGTGTCT-----AGCATAGAT---GTTTCTCATCAAT 1638  
 Db 656 AsnAsnPheIleSerLysProLeuArgAlaAsnAsnArgValGluLysLeuArgLysAsp 675  
 QY 1639 GAGTGTATGATGGTCTCTATGCAGAAATCTCAGGATATTTCATCGCAACCTGGATTT 1698  
 Db 676 GluLeuTyrAsnGlySer-----LysGlnLysArgArgThrArgGlySer 690  
 QY 1699 GCTTTTATGTCGCAACGACCACTTAAG----- 1725  
 Db 691 AlaProAspSerGluGlySerLysLysProLeuIleGluArgGlnIleGluAspGlyAsn 710  
 QY 1726 -----AGAACACTATCACACATATTCTGTGTGCGAGCGCAACAAACGA 1767  
 Db 711 ValSerGluLysThrProGlnGluProThrThrSerIle-----ProGlyGlu 726  
 QY 1768 ACAAAAGCAAGCATGCTGTAATTTCTTGAATCTGAAGTGGGGAAGTGAACAGCAAGA 1827  
 Db 727 ThrAspValAsnGluLysValLeuGluProIleGluAsp-----IleGluMetLeuPro 744  
 QY 1828 ACATATAGTAGGCGCCACAACTCTGCTGATTTTCCATAGTGATACC-----TGC 1875  
 Db 745 ThrPheIleSerSerSerAsnArgLeuGluProAspAlaAsnThrLeuSerAlaHisAla 764  
 QY 1876 TTACCTT----- 1881  
 Db 765 LeuProSerSerValIleGlyValLysLysGlnProSerLeuThrThrValAsnLysPhe 784  
 QY 1882 -----CTCCGTCCACAGAAATGGAAGTAGATAGTGA--- 1914  
 Db 785 ProAlaSerPheGluAsnGlyLeuGlnProAspAsnAlaGlnAlaSerThrGluIleGlu 804  
 QY 1915 ---GATGAAGAAGATCCTGAATGGCTTAAGAGAAAAAACCAATTACACAAATTTGAAGATTT 1971  
 Db 805 ThrAsnGluGlyAsnSerLysLysGlyLysAspGluLeuHisPheAspProGluLysCys 824  
 QY 1972 TCTGATGTTAAGAGGAGAGAAAGATGATGAACCTCTCGAATCTCCATGTCATGAAG 2031  
 Db 825 GluAspValThrGlnPheArgLysGluTyrLeuAlaLeu---ArgAlaSerIleSerLys 843  
 QY 2032 CATGGGTTTTATTGCTCACAAATCAATCATGCTGTATGCTGTTTGTAGAAAAATTAT 2091



```

Db      844 AspAspThrCysSerSerLeuValAsn----- 853
QY      2092 GGACAGAAATATTAGAGAAATTTATGTCGAACTTCATGCTTCATCTAGTCACGATG 2151
Db      854 ---GluAspLeuAsnLysThrGluValIleGlnAsnLeuThrSerGluIleIle---Gln 871
QY      2152 CATGACTTTAATCTATTAGCATAATGTCAATAGATAAAGCTGTTACCAAGCTCCGCTGAA 2211
Db      872 SerGluMetLysIleAlaSerLeuValGlyIleAsnHis-----GlnLeuArgGlu 888
QY      2212 ATGCACCAAAATAGAAAG-----GGGGAATCTGCTTCCCTGCAACGAAAGATA 2265
Db      889 LysLeuGluGluGluGluLysIleGlnAlaLysLeuValGlnGluLeuAspThrValLeu 908
QY      2266 ACTGAGAAACAAAT-----GGGACAGCAATGGATTAGTGAATTAACATCA 2313
Db      909 LeuGluSerLysGlySerPheSerAlaGlyThrAlaLysGlu---AlaGluValAsnAsn 927
QY      2314 AAAGAGAAAGCTTTGGAACACAGATAGTGTCTCAGGGGTTTCAAAACAGACAAAACAA 2373
Db      928 LysAspLys-----ThrGluSerIleAsn-----AsnGluGlnAlaArgGlu--- 941
QY      2374 AAATCTGAAAAGCTCTAACCCCATGTTATGGACAAACACTG 2415
Db      942 -----LysSerLeuThrAlaAlaHisGlyLysLeuIle 952

RESULT 3
S59359
Gln4 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YDR507c
C:Species: Saccharomyces cerevisiae
C:Date: 30-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 24-Sep-1999
A:Accession: S59359; S69565
R:longline, M.S.; Fringle, J.R.
submitted to the EMBL Data Library, August 1995
A:Reference number: S59359
A:Molecule type: DNA
A:Accession: S59359
A:Residues: 1-1142 <LON>
A:Cross-references: EMBL:U33140; NID:g992650; PIDN:AAA75513.1; PID:g992651
R:Dieckrich, F.S.
submitted to the EMBL Data Library, August 1995
A:Description: The sequence of S. cerevisiae cosmid 81.66, 9787, 9717, and lambda 3073.
A:Reference number: S69553
A:Accession: S69565
A:Molecule type: DNA
A:Residues: 1-1142 <DIS>
A:Cross-references: EMBL:U33057; NID:g927764; PIDN:AAB64949.1; PID:g927777; MIPS:YDR507C
C:Genetics:
A:Gene: SGD:GIN4
A:Cross-references: SGD:S0002915; MIPS:YDR507C
A:Map position: 4R
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C:Keywords: ATP
F:17-289/Domain: protein kinase homology <KIN>
F:25-33/Region: protein kinase ATP-binding motif

Alignment Scores:
Pred. No.: 7,026-05 Length: 1142
Score: 176.50 Matches: 152
Percent Similarity: 34.75% Conservative: 102
Best Local Similarity: 20.79% Mismatches: 264
Query Match: 2.30% Indels: 213
DB: 2 Gaps: 35

US-09-874-162A-7 (1-4409) x S59359 (1-1142)
QY      172 GATCCACGGGTTTATAGAAAACAGAAATTACAGACGCA-----ACC 213
Db      337 AspProGluGlyIle---LysGluLysLeuArgGluProGlyAlaAsnAlaGluLysThr 355
QY      214 TATGTTCCTGAGTTACATAAATAGATTTCACAGATGTCGCCGCCGACGACGAGAG 273

```

```

Db      356 LeuTyralaLeuLeuTyrr-----ArgPheLysCysAspThr-----GlnLys 369
QY      274 TCCCTTAAGAGAGAGATTACGCCAAGCTTCGCTGACTCTGTCCAGC---TCAGTGTCT 330
Db      370 GluLeuIleLysGlnGlnGlnValLysArgGlnSerIleSerValSerValSer 389
QY      331 CGAGGG---AATGTGTCCACTCCCCCA-----CGCCACACAGTGGAGAC 372
Db      390 ProSerLysLysValSerThrThrProGlnArgArgAsnArgGluSerLeuLys 409
QY      373 CTTACTCCCCCGTCAGCCCCACCCCATC-----ACCCCTCCTCT 411
Db      410 ValThrSerSerArgLysLysProIleSerPheAsnLysPheThrAlaSerSerAlaSer 429
QY      412 TCATTCCGCACGACACTCCGACAGACCCACAGATCTATAGATTCTTCGNACTCGG 471
Db      430 SerSerAsnLeuThrThrProGlySerSerLysArgLeuSerLysAsnPheSerSerLys 449
QY      472 AATCTCATAGCACCACATATTTTTGCACAGAACTTCTTACTTACATGCTCATCGAACTCC 531
Db      450 LysLysLeuSerThrIleValAsnGlnSerSerProThrProAlaSer---ArgAsnLys 468
QY      532 AGAACAAACATC-----AAAGAGAAACATTTAAAGTTCATGATATGTTATCAAAAGTA 585
Db      469 ArgAlaSerValIleAsnValGluLysAsnGlnLysArgAlaSerIlePheSerThr 488
QY      586 GAGAAATGAAGAGAGACAGCAAGATCT----- 612
Db      489 LysLysAsnLysArgSerSerArgSerIleLysArgMetSerLeuIleProSerMetLys 508
QY      613 ---CATAGCTGTGACGCTCATTTGCAGCTTACGTTTACTGTTTCTTCCACAAATGAT 669
Db      509 ArgGluSerValThrThrLysLeuMetSerThrTyralaLysLeuAlaGluAspAsp 528
QY      670 AGCCATCACCACAACTCAGAAATGAACAAATCTGTACCTCCCTGGAGTCCCTCTGTG 729
Db      529 TrpGluTyrlleGluLysGluThrLysArgThrSerSerAsnPheAlaThrLeuIleAsp 548
QY      730 AAAGTT-----TGCCACAAAAAAGAGAGATGTAAGTTGT 765
Db      549 GluIlePheGluTyrrGluLysTyrrGluGlnIleArgLysGluLysGluGluArg 568
QY      766 CCAATAGGCAAGTCCACAGGT----- 789
Db      569 LysValArgGluAlaLysAlaArgGluGluArgArgArgLysGlnGluGlu 588
QY      789 ----- 789
Db      589 LysGluArgAlaArgLysLeuLeuLysGluAspLysArgLysGlnGluLeu 608
QY      790 AAAAAGCAGTGCCTTTG---ATTCCTGACCTCAATCAA----- 825
Db      609 LysLysGlnIleGluIleAspIleSerAspLeuGluGlnGluLeuSerLysHisLysGlu 628
QY      826 ---ACAAAACCCGAAATTTCCCGTCCCTGAGTTTCCAGTAATCAATTTGACCTAGT 882
Db      629 GluLysLeuAspGlyAsnIleArgSerIleSerAlaProMetGluAsnGluGluAsn 648
QY      883 AACAGCCATATGTTGAGAGTCTTACTCGTTGCTATTAGA-----GTG 924
Db      649 IleAsnHisLeuGluValAspIleAspAsnIleLeuArgArgArgAsnPheSerLeuGln 668
QY      925 ACTGTCGACGAGAGAGAGAGTTTAATGATGATTAATGAGAAACCAATGAAATATT 984
Db      669 ThrArgProValSerArgLeuAspProGlyIleMetPheSerSerProThrGluGluVal 688
QY      985 GATGTCATGAGAGCTTCAGCCAGAGAAATCGTGAGATCGGAGAGACAAAGACA 1044
Db      689 SerProValGlu-----ProLysArgThrGluAsnGluArgLeuThrThrGluLys 706
QY      1045 TTTGTTGCACAAATG-----ACAGTATTTGATAAAACAGGCGC 1083

```

Db 707 IleLeuGluThrIleArgArgSerLysPheLeuGlySerSerPheAsnIleAspLysGlu 726  
 QY 1084 TTACAGCTTTTAGATGGGAATATGAA-----GTAGCCATGCAGGAATG--GAAGAA 1134  
 Db 727 LeuLysLeuSerLysMetGluTyProSerIleIleAlaProGlnArgLeuSerGluGlu 746  
 QY 1135 TGTCGAATAGCAAGAAAGACACATGGAGACTATTCTT-----GATGGGAAG 1185  
 Db 747 ArgValValSerSerAsnAspGlyTyProGluSerLeuIleLeuProLysAspGlyAsn 766  
 QY 1186 AGGCTGCCTCCATTCGAA-----ACATTTTCTCAGGACCTAGCTTGCAGTTCACTCTT 1239  
 Db 767 GlyValSerGlnLeuLysAspSerThrAlaThrAlaProValSerAspGlyArgLeu 786  
 QY 1240 CTTGGACAGAGAGACCAATGATTAATCTAGGCTCTTATGCAACCTCTTGGCACT 1299  
 Db 787 ArgLysIleSerGlu-----IleArgValProGlnPheThr 798  
 QY 1300 AGAAATTCAGAGAGCTCCATCAGGAAACAGCTGTTTCAGTTTAAACCTTAACT 1359  
 Db 799 ArgLysSerArgHisPheSerGluSerAsnLysArgLeuSerVal-----LeuSerMet 816  
 QY 1360 ATTGCTGTTAAAGATCAATGACTACAGATCTCAAAACAGAAAGAAAGATACCTCCA 1419  
 Db 817 TyrSerThrLysGluSerPheThrAsnLeuValAspIleLeuLysAsnGlyAsnLeuAsp 836  
 QY 1420 AATGAAACCGCAAAATTAAGATATTTTATCAGTTTCTTATACAAATACAAGG 1479  
 Db 837 ValAsnAsnGlnSerGlnArgIle----- 845  
 QY 1480 CAACAACTGAAGCAAGAGACCTGCAATGCCCTTGGTGTACTCTGAATGCCGCAAA 1539  
 Db 846 ProThrProArgSerAlaAspAsp----- 853  
 QY 1540 CTTTATAGTTTACTCAAGCACTTAACTCTGCCATACAGATTTATCTTCACTATGTT 1599  
 Db 854 -----SerGluPheLeuPheGlu----- 859  
 QY 1600 TATCATCAAAAGGTGCTAGGATAGATGTTTCTATCAATGAGTGTATGATGGCTCTAT 1659  
 Db 860 -----ThrValAsnGlu-----GluAlaGluTyr 867  
 QY 1660 GCAGGAATCTTCAGATATTCATCGC----- 1686  
 Db 868 ThrGlyAsnSerSerAsnAspGluArgLeuTyraPheValGlyAspSerThrIleLysAsp 887  
 QY 1687 -----CAACTGGATTGCT-----TTTACTCGAACGGACCACTTAAGAGA 1728  
 Db 888 LysSerAlaLeuLysLeuAsnPheAlaAspArgPheAsnGlySerAsnGluAlaGln 907  
 QY 1729 ACACCTATCACATATTTCTGTGTGCGGCAAAACGAAACAAAGCAAGCATGCTGAA 1788  
 Db 908 ThrAspAsnLeuHisLeuProIleLeuProLeuAsnGlyAspAsnGluLeuArgLys 927  
 QY 1789 TTTCTTGAATCTGAAGATGGGAAGTAGAACAGCAAGAAACATATAGT----- 1836  
 Db 928 ---GlnAsnSerGlnGluLysAspGlnAlaHisProLysIleLysSerMetIleProGlu 946  
 QY 1837 AGTGGCCAACTCGTGTATTTCCATAGTATACCTGCTTACTCTCCGTCCTCAAGAA 1896  
 Db 947 SerGlySerSerSer-----HisThrGlu-----LysGlu 956  
 QY 1897 ATGGAAGTAGATGAGATGAAGATCAAGATCTCTGAATG----- 1935  
 Db 957 GluGluAsnGluGluLysGluLysLysProGluGlnHisLysGlnGluAspGln 976  
 QY 1936 ---CTAAGAGAAACCAATACACAAATTGAA 1965  
 Db 977 GluLysArgGluLysValValAspAspMetGlu 987

RESULT 4

T42631

breast cancer tumor suppressor BRCA2 - dog (fragment)

C;Species: Canis lupus familiaris (dog)  
 C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 24-Nov-2003  
 C;Accession: T42631  
 R;Bignell, G.R.; Mickleth, G.; Stratton, M.R.; Ashworth, A.; Wooster, R.  
 submitted to the EMBL Data Library, July 1996  
 A;Description: BRCA2 contains a novel repeat that is conserved in other mammian species.  
 A;Reference number: Z22184  
 A;Accession: T42631  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1507 <BIG>  
 A;Cross-references: EMBL:275664; NID:e936328; PID:e251256; PIDN:CAA99994.1  
 C;Genetics:  
 A;Gene: BRCA2  
 C;Superfamily: breast/ovarian cancer tumor suppressor protein, BRCA2 type  
 C;Keywords: tumor suppressor

## Alignment Scores:

Pred. No.:	0.000431	Length:	1507
Score:	166.00	Matches:	180
Percent Similarity:	33.96%	Conservative:	143
Best Local Similarity:	18.93%	Mismatches:	384
Query Match:	2.16%	Indels:	244
DB:	2	Gaps:	43

US-09-874-162a-7 (1-4409) x T42631 (1-1507)

QY 118 TTCCCCAGCCTCGCGACCTCATCGAGCACATCGAGACCAACCATCGATACAGATCCA 177  
 Db 373 PheLysThrAlaSerAsnLysLysLysLysLysLysLysLysLysLysLysLysLys 392  
 QY 178 CGGGTTTGAAGAAACAGATTCAGAGAGCAACATGTTGTCCTGAGGAGATGTTCCACTCCCCA 354  
 Db 393 MetLeuPheLysAspIleGluHisTyrProThrAsnLeuAlaCysLeuGluIleVal 412  
 QY 235 AATAGATTCATCAGATCTGCTCCGCGAGAGAGGAGTCCCTAAAGAAAGAGATTCAG 294  
 Db 413 AsnThrSerSerLeuGluSerGlnLysLysProSerLysSer-----HisAlaLeuAsp 430  
 QY 295 CGCAAGCTCTCGTCACTCTGTCAGCTCAGTGTCTCGAGGAGATGTTCCACTCCCCA 354  
 Db 431 ProGlnSerIleAsnIleSerGlyPheValGln-----AsnSerThrTyrValSer 448  
 QY 355 CGCCACAGAGTGGAGAGCTTACTCCCGCGTAGCCACCCACCATCACCCTCTCTTCA 414  
 Db 449 AspSerGluSerGlyHisThrAlaProThrLeuSerLeuLysGlnAspPheAspSer 468  
 QY 415 TTCCGCGAGCAGCACTCCGACAGAGCAACACACAGATC----- 450  
 Db 469 AsnArgAsnLeuThrProSerGlnLysAlaGluIleThrGluLeuSerThrIleLeuGlu 488  
 QY 451 -----TATAGATTTCTTCACTCGGAATCTCATAGCACAATATTTTG 495  
 Db 489 GluSerGlySerGlnPheGluPheThrGlnPheArgLys-----ProSerHisIle 505  
 QY 496 CACAGAACTTCTTACTTACATGCTCATCGAAATCCAGAACCA-----AACATCAAAAG 549  
 Db 506 IleGlnLysAsnProPheGluMetProGluAsnGlnLeuThrIleLeuAsnSerThrSer 525  
 QY 550 AAAACATTTAAAGTTCATGATATG-----GGAGACAGAGATCTCATAGCTTGTGACGT 627  
 Db 526 LysGluTyrLysAspAspAspLeuHisLeuThrAsnAlaProSerIleSerGlnVal 545  
 QY 586 GAGAAATGAA----- 585  
 Db 546 AspSerLysLysSerGluGlyIleIleGlyGlyGlnLysPheAlaCysLeuSerArg 565  
 QY 628 -----CATTTGCAGCTTACGTTTACT 648  
 Db 566 ThrSerCysAsnArgSerAlaSerGlyTyrSerThrAspLysAsnGluValGluPheArg 585  
 QY 649 GGTTCCTTCCAC---AAAAATGATAGCCATCACCAACTCAGAA----- 690

Db 586 GlyPheTyrSerAlaArgGlyThrLysLeuAsnValGlySerGluAlaLeuGlnLysAla 605  
QY 691 -----ATGAAACAAATTCGTTACCTCGAAGTC 720  
Db 606 LysLysLeuPheSerAspLeuGluAsnIleAsnGluGluThrSerValGluValAsp--- 624  
QY 721 CTGCTTGTAAGTTGGCCACAAAAAGAAAGAGTAAAGTTGTCTCAATAAGCAAGTT 780  
Db 625 -----ArgSerPheSerSerSerLysTyrAsnAsp 634  
QY 781 CCCACAGTAAAGCAGTGCCTTTGATCTCGACCTCAATCAACAAACCCGAAAT 840  
Db 635 SerValSerMetIleGlnIleGluAspCysAsnAspLysAsnLeuAsnGluProAsnAsn 654  
QY 841 TTCCTGCTCCCTGACGATCCAGTATGATTAATTTGAACCTAGTAACAGCCATATGTGAG 900  
Db 655 LysCysArgLeu---IleLeuGlnAsnAsnIleGluMetThrThrAspIlePheValGlu 673  
QY 901 TCTTACTCTGCTGCTATTTAGA-----GTGACTCGTCCA 933  
Db 674 GluTyrThrGluSerTyrArgArgAsnThrGluAsnGluGlyAsnGlnCysThrAspAla 693  
QY 934 GGAAGAGAGAGTTTAATCGAATGATTAATGGA-----GAAACCAATGAAATATT 984  
Db 694 GlyArgAsnThrCysAsnSerGluSerAspGlySerAspSerSerLysAsnAspThrVal 713  
QY 985 GATGTCATCAATGAAGAG-----CTTCCAGCCAGAAAGAAATCGTGAGGATGGG 1035  
Db 714 TyrIleHisGluGluGluAsnGlyLeuProCysIleAspGlnHisAsnIleAsp----- 731  
QY 1036 GAAAGACATTTGTCGCAAAATGACAGTATTTGATTAATAAAACAGCGCTTACAGCTTTTA 1095  
Db 732 -----LeuLysLeuPheSerGlnPheMetLysGluGlyAsnThrGlnIleLys 747  
QY 1096 GATGGGAATATGAAGTA-----GCCATGCAGGAATCGAAGAATGCCAATA 1143  
Db 748 GluGlyLeuSerAspLeuThrCysLeuGluValMetLysAlaGluGluThrSerHisVal 767  
QY 1144 ACCAAGAAAAGCAACATCGGAGAGATTTCTTGATGGGAAGAGCGCTCCCTCCATTCGAA 1203  
Db 768 ThrMetSerAsnLysGlnLeuThrAlaAsnThrGlyGlnAsnIleLysAspPheAsp 787  
QY 1204 ACATTT-----TCTCAGGAGCTACGTTGCGAGTTCACTCTCTGTTGACAGGAGAGACC 1257  
Db 788 ThrPheTyrLeuSerPheGlnThrAlaSerArgLysAsnIleArgValSerArgGluSer 807  
QY 1258 AATGATAATCTACGCGCTTATTCGCAACCTCTTGCACATAGAAAT----- 1305  
Db 808 LeuAsnLysAlaArgSerLeuLeuAsnGlnLysTrpThrGluGluLeuAsnAsnPhe 827  
QY 1306 TCAGAGAGTCCCATCAGGAAACAAAGCTGTG---TCAGTTAAACCTACTCAAACTATT 1362  
Db 828 SerAspSerLeuAsnSerGluLeuLeuProGlyIleAspIleLysLysThrAsp----- 845  
QY 1363 GCTGTTAAGAATCATCTGACTACAGATCTACAAACAGAAAGAAAGATCTCCAAAT 1422  
Db 846 ---IleSerAsnHisGluValIleGluAsnThrGluArgLysAspLysIleThrLysGlu 864  
QY 1423 GAAACCCGACAAAAATTAAGAAATATTTTATCAGTTTCTCTATAACAAATACAAAGCAA 1482  
Db 865 SerAspLeuIleGlyThrGluAsnIleLeuLeuIleLeuGlnArgProGluSerLys 884  
QY 1483 CAAACTGACGACAGAGATGACCTGCATTGCCCTTGGTGTACTCTGAACCTGCCGCAACTT 1542  
Db 885 IleLysLysIleLysGluSerAlaValLeuGlyPheHisThrAlaSerGlyLysLysIle 904  
QY 1543 TATAGTTTACTCAACATCTTAAACTCTGCCATACGATTT----- 1584  
Db 905 GluIleThrLysGluSerLeuAspLysValLysAsnLeuPheGluGluLysGluGluAsp 924  
QY 1585 -----ATCTCAACTATGTTTATCATCCAAAAGGTGTAGATAGATGTTTCTATC 1635  
Db 925 AsnSerGluIleThrAsnPheSerHis-----ArgGlyAlaLysMetSerLysAspArg 942

QY 1636 AATCAGTGTATGATGC---TCCTATGCAGAAATCCTCAGGATATTCATCGCCAACCT 1692  
Db 943 GluGlyCysLysAspGlyArgGluLeuAlaCysGlyThrThrGluIleThrThrPro 962  
QY 1693 GGATTTCGTTTACTCCCAACGACAGTAAAGAGAACACCTATCACACAT---ATTCTT 1749  
Db 963 GluTyrGluGluThrHisSerSerLeuGluLysLysLysLeuValSerAsnGluIleAla 982  
QY 1750 GTGTGCGAGCCA-----AAACGAAACA----- 1770  
Db 983 AlaLeuArgProArgLeuLeuSerAspAsnLeuTyrLysGlnThrGluAsnLeuLysIle 1002  
QY 1771 -----AAACGACGATGCTGTAATTTCTTGAATCTGAATCGAAGTGGGAA 1812  
Db 1003 SerAspHisAlaSerGlnLysValAspValHisGluAsnThrGluLysGluThrAlaLys 1022  
QY 1813 -----GTGAACACAGAAAGAACATATAGTAGTGGCCACCAATCGT---CTCTAT 1857  
Db 1023 LysProThrMetTyrThrAsnGlnSerThrTyrSerAlaIleGluAsnSerProLeuThr 1042  
QY 1858 TTCCATAGTATACCTGCTTACCTCTCCGTCACAGAAATGGAAGTAGTAGTGAAGAT 1917  
Db 1043 PheThrGlnAspThr-----GluGluLysPheSerValSerGluAla 1056  
QY 1918 GAA-----AAGGATCCTGATCGCTAAGAGAAACCACTTACACAAATGAAGATT 1971  
Db 1057 SerLeuPheGluAlaLysLysTrpLeuArgGlu-----GlyGluTrp 1070  
QY 1972 TCTGAT-----GTTAATGAGAGAGAGAGAAAGATGATGAACTCTGG----- 2013  
Db 1071 AspAspGlnSerGluArgIleAsnAlaAlaLysValAsnCysLysLysGluTyrProAsp 1090  
QY 2014 -----AATCTCCATGTCATGACATCGGTTTATTGCTGCAATCAA 2055  
Db 1091 AspTyrValGluAsnProSerCysGlyAsnSerSerAsnSerAlaIleThrGluAsnAsp 1110  
QY 2056 ATGAATCATGCTGATGCTGTTGTAGAAAATTTATGGACAGAAAATAATTAAGAAAT 2115  
Db 1111 LysAsnHis-----LeuSerGluLysGlnGlySerThrTyrLeuSerAsnSer 1126  
QY 2116 TTATGTCGAAAC---TTCATGCTTCTCTAGTCAGCATGCAATGAC-----TTTAAT 2163  
Db 1127 ThrMetSerAsnSerTyrSerTyrHisProGlyPheCysHisSerSerGluValTyrAsn 1146  
QY 2164 CTTATTAGCATAATGTCA-----ATAGATAAAGCTGTACC 2199  
Db 1147 LysSerGluTyrLeuSerArgSerLysIleAspAsnSerGlyIleGluProValIleLys 1166  
QY 2200 AAGCTCCGTGAAATGCAG----- 2217  
Db 1167 AsnIleArgGluArgLysAsnIleGlyPheSerGluIleMetSerProGlyArgGluAla 1186  
QY 2218 -----CAAAAATTAGAAAGGGGAA 2238  
Db 1187 AspThrAspProGlnSerValAsnGluAspIleCysValGluLysLeuAlaThrAsnSer 1206  
QY 2239 TCTGCTTCCCTCCAAAC----- 2256  
Db 1207 SerCysLysAsnLysAsnThrAlaIleLysValAlaIleSerAspSerAsnAsnPheAsn 1226  
QY 2257 -----GAAGAAATACTGAAGAACAAAATGGGACAGCAAAATGGATTTAGTGAATAAC 2310  
Db 1227 ThrIleGlnLysLeuAsnSerAspSerAsnAsnSerValProAlaTyrSerThrValAsn 1246  
QY 2311 TCAAAAGAG-----AAGCTTTGGAA-----ACAGATATGTGC 2343  
Db 1247 SerLysArgValPheValAlaHisGlnThrLysValThrGluGlyPheThrAspAsnCys 1266  
QY 2344 TCAGGGGTTTCAAAACAGAGCAAAAAACAAAAA 2376  
Db 1267 SerMetValThrLysGlnAsnThrLysSerLys 1277

## RESULT 5

S38173  
 myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YKR095w; protein YKR415  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 03-May-1994 #sequence revision 03-May-1994 #text\_change 21-Jul-2000  
 C:Accession: S38173; S40647; S31207  
 R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, submitted to the Protein Sequence Database, March 1994  
 A:Reference number: S38158  
 A:Accession: S38173  
 A:Molecule type: DNA  
 A:Residues: 1-1875 <BAL>  
 A:Cross-references: EMBL:Z28320; NID:5486586; PID:5486587; MIPS:YKR095w  
 A:Experimental source: strain S288C  
 R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jil Yeast 9, 1349-1354, 1993  
 A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chrom  
 A:Reference number: S40644; MUID:94205265; PMID:8154186  
 A:Accession: S40647  
 A:Molecule type: DNA  
 A:Residues: 1-1875 <BOU>  
 A:Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554  
 A:Experimental source: strain S288C  
 R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.  
 Mol. Gen. Genet. 237, 359-369, 1993  
 A:Title: A new yeast gene with a myosin-like heptad repeat structure.  
 A:Reference number: S31207; MUID:93247549; PMID:8483450  
 A:Accession: S31207  
 A:Molecule type: DNA  
 A:Residues: 1-300, 'A', 302-1875 <KOE>  
 A:Cross-references: EMBL:L01992; NID:G171958; PIDN:AAA34783.1; PID:G171959  
 C:Genetics:  
 A:Gene: SGD:MLP1  
 A:Cross-references: SGD:S0001803; MIPS:YKR095w  
 A:Map position: 11R

Alignment Scores:  
 Pred. No.: 0.00103 Length: 1875  
 Score: 161.00 Matches: 167  
 Percent Similarity: 36.63% Conservative: 129  
 Best Local Similarity: 20.67% Mismatches: 297  
 Query Match: 2.09% Indels: 216  
 DB: 2 Gaps: 35

US-09-874-162A-7 (1-4409) x S38173 (1-1875)

QY	410	CTTCATTCGCGAGCAGCAGCTC---CGACAGACCAACACAGATCTATAGATTCTTCGAA	466
Db	191	LeuAsnSerArgThrGluLeuGluArgLysThrGlnGluLeuThrLeu-LeuGlnSerAs	210
QY	467	CTCGGAATCTCATACACCAATATTTTTCGACAGACTCTTACTTACATGTCTCATCGA	525
Db	210	nAsnAspTrpLeuGluLysGluLeuArgSerLysAsnGlu-GlnTyrLeuSerTyrArgG	230
QY	526	-----AACTCCAGACAAACATCAAAAGGAAACAT	556
Db	230	lnLysThrAspLysValIleLeuAspIleArgAsnGlnLeuAsnArgLeuArgAsnAsp	250
QY	557	TTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGACAGCAAGTCTCAT	616
Db	250	heGlnMetGluArgThrAsnAsnAspValLeuLysGlnLysAsnGlnLeuLeuSerLys	270
QY	617	GTTTGTACGCTCATTTGAGCTTACGTTTACTGGTTTCTTCCACAAAATGATGACCAT	676
Db	270	erLeuGlnGluLysLeu---LeuGluIleLysGlyLeu-----SerAspS	284
QY	677	CACCAATCTCAGAAATGAACAAAATCTGTGTACCTCGAAGTCTGCTGTCTGAAAGTTT	736
Db	284	erLeuAsnSerGluLysGlnGluPheSerAlaGluMetSerLeu-----	298
QY	737	GCCACAAAAGAAAGAGTGTAAAGTTGTCCATTAAGCAAGTCCACACAGTAAAGAC	796

Db	299	-----LysGlnArgLeuValAspLeuLeuGluSerGlnLeuAsnAlaValLysGluG	316
QY	797	AGGTGCTTTGATTCCTGCACCTCAATCAAAACAAACCCGGAATTTCCGCTCCCTTGAG	856
Db	316	luLeuAsnSerIleArgGluLeuAsnThrAlaLys-----ValIleAlaA	331
QY	857	TTTCCAGTAATGAATTTGAACCTAGTACACGCCATATGTTGAAGTCTTACTGCTGTAT	916
Db	331	spAspSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys	351
QY	917	TTAGAGTGACTCGTCCAGGAACAGAGAGTTTAAATGAATGATTAATCGGAACCAATG	976
Db	351	ysGluLysLeuAlaGlnCysGluLysGluCysLeu---ArgLeuSerSerIleThrAspG	370
QY	977	AAAAATTGATTCATCAATGAAGAGCTTCAGCC	1008
Db	370	luAlaAspGluAspAsnGluAsnLeuSerAlaLysSerSerSerAspPheIlePheLeuL	390
QY	1009	-----AGAAAGAAACGAAATCCTGAGGATGGGAAAGACAT	1045
Db	390	ysLysGlnLeuLysGluArgArgThrLysGluHisLeuGlnAsnGlnLeuGluThrP	410
QY	1046	TTGTTGCACAAATG-----ACAGTATTTGATAAAAAACAGGCGCTTACAGC	1090
Db	410	heIleValGluLeuGluHisLysValProIleAsnSerPheLysGluArgThrAspM	430
QY	1091	TTTTAGATGGGAATATGAAGTAGCCATGCAGGAAATGGAAGATGTCCAATAGCACA	1150
Db	430	etLeuGluAsnGluLeuAsnAsnAlaLeuLeuLeuGluHisThrSerAsnGluLysA	450
QY	1151	AAAGAGCAATGGGAGACTATTTCTGATGGAGAGCTGCTCCATTCGAAACATTTT	1210
Db	450	salalysValLysGluLeuAsnAlaLysAsnGlnLysLeuValGluCysGluAsnAspL	470
QY	1211	CTCAGGACCTACG-----TTGAGTTCCTACTCTTCGTT	1243
Db	470	euGlnThrLeuThrLysGlnArgLeuAspLeuLysCysArgGlnIleGlnIleThrLeuL	490
QY	1244	GGACAGGAGAGACCAATGATAATCTACGGCTCCTATTGCCAAA	1287
Db	490	hrAsnSerValSerAsnAsp---SerLysGlyProLeuArgLysGluIleGlnPheI	509
QY	1287	-----	1287
Db	509	leGlnAsnIleMetGlnGluAspAspSerThrIleThrGluSerAspSerGlnLysVal	529
QY	1288	-----CCTCTTCCACTAGAAATTC	1308
Db	529	aiThrGluArgLeuValGluPheLysAsnIleGlnLeuGlnGlnLysAsnAlaGluL	549
QY	1309	-----GAGAGTCTCCATCAGGAAAAACAAGC	1333
Db	549	eulLysValValArgAsnLeuAlaAspLysLeuGluSerLysGluLysLysSerLysG	569
QY	1334	CTGGTTACGTTAAA---CCTACTCAAACTATT---GCTGTTAAAGATCATTTGACTACA	1386
Db	569	lnSerLeuGlnLysIleGluSerGluThrValAsnGluAlaLysGluAlaIleThrL	589
QY	1387	-----GATCTACAAACAAAGAAAAG	1410
Db	589	eulysSerGluLysMetAspLeuGluSerArgIleGluLeuGlnLysGluLeuGluG	609
QY	1411	-----GATACCTCCAAATGAAACCCGACAAATAAAGATATTTTATCATGTTTC	1459
Db	609	luLeuLysThrSerValProAsnGluAspAlaSer-----	620
QY	1460	TCTATACACACATACA---AGCACAACCTGACCAAGACATGACCTGCATTGCCCTT	1516
Db	621	--TyrSerAsnValThrIleLysGlnLeuThrGluThrLysArgAspLeuGluSerGlnV	640
QY	1517	GGTGTACTCTGAAC-----TGCCGCAAACTTTATAGTTTACTCAAGC	1558
Db	640	alGlnAspLeuGlnThrArgIleSerGlnIleThrArgGluSerThrGluAsnMet-Ser	659

1559 ATCTTAAACTCTGCCATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTA 1618  
 Db :|||||  
 660 LeuLeuAsnLysGluLeuGlnAspLeuTyrAspSerLysSerAspLysSerLeuLysLeu 679  
 Qy :|||||  
 1619 GGTATAGATGTTCTCAATAGTGTATGATGGCTCTCTATGCGGAAATCTCTCAGGATA 1678  
 Db :|||||  
 680 Gly-----LysGluLysSerSerArgIleLeuAla-----GluGluArgPheLysLeu 695  
 Qy :|||||  
 1679 TTCATCGCCCACTGATGTTGCTTTTAGTCGCAACGGACAGCATGTTAAGAGAACACCTATCA 1738  
 Db :|||||  
 696 LeuSerAsnThrLeuAspLeuThrLysAlaGluAsnAspGlnLeuArgLysArgPheAsp 715  
 Qy :|||||  
 1739 CACATATT-----CTTGTGTGCGGCCCAACCAACCAACCAACCAACCAACCAACCAACCA 1791  
 Db :|||||  
 716 TyrLeuGlnAsnThrIleLeuLysGlnAspSerLysThrHisGluThrLeuAsnGluLys 735  
 Qy :|||||  
 1792 CTT-----GAATCT 1800  
 Db :|||  
 736 ValSerCysLysSerLysLeuSerIleValGluThrGluLeuLeuAsnLeuLysGluGlu 755  
 Qy :|||  
 1801 GAAGATGGGGAAGTAGAACAGCAAGAACATATAGTAGTGGCCACCAATCGTCTGTATTTC 1860  
 Db :|||  
 756 GlnLysLeuArgValHisLeuGluLysAsnLeuLysGlnGluLeuAsnLysLeuSerPro 775  
 Qy :|||  
 1861 CATAGTACTCTGCTTACTCTCTCGTCCACAGAAATGGAAGTAGATAGTAGAAGATGAA 1920  
 Db :|||  
 776 GluLysAspSerLeuArgIleMetValThrGlnLeuGlnThrLeuGlnLysGluArgGlu 795  
 Qy :|||  
 1921 AAGGATCCTGAATGGCTAAGAGAAAACCAATACACAAATGAAGAGTTTCTGAT--- 1977  
 Db :|||  
 796 AspLeuLeuGluGluThrArgLysSerCysGlnLysLysIleAspGluLeuGluAspAla 815  
 Qy :|||  
 1978 GTTAATGAAGAGAGAAAGTATGAACTCTGGAATCTCCATGTCATGAAGCATGGG 2037  
 Db :|||  
 816 LeuSerGluLeuLysLysGluThrSerGln-----LysAspHisIleLysGln--- 832  
 Qy :|||  
 2038 TTATGCTGCAATCAATGAATCATGCTGCTGATGCTGTTGTAGAAAATTAATGGACAG 2097  
 Db :|||  
 833 LeuGluGluAspAsnSerAsn-----IleGluTyrTyrGlnAsn 846  
 Qy :|||  
 2098 AAAATA-----ATTAAGAGAAATTAATGTCGAAACTTCATGCTTCATGTCAGCATG 2151  
 Db :|||  
 847 LysIleGluAlaLeuLysLys----- 853  
 Qy :|||  
 2152 CATGACTTTAATCTTATTAGCATATATGTAATAGATAAAGCTGTTTACCAGTCCGTGAA 2211  
 Db :|||  
 854 ---AspTyrGluSerVal---IleThrSerValAspSerLysGlnThrAspIleGluLys 871  
 Qy :|||  
 2212 ATGCAGCAAAATTAAGAAAGGGGAATCTGCTTCCCTGCAACGACGAGAAATAACTGAA 2271  
 Db :|||  
 872 LeuGlnTyrLysValLysSerLeuGlu-----LysGluIleGluGlu 885  
 Qy :|||  
 2272 GAACAAAATGGGACAGCAAAATGGATTATGTAATAACTCAAAAGAGAAAGCTTTGGAA 2331  
 Db :|||  
 886 AspLys-----IleArgLeuHisThrTyrAsnValMetAspGluThrIleAsn 901  
 Qy :|||  
 2332 ACAGATAGTGTCTCAGGGGTTTCAAAACAGACGACAAAACAAACTCTGAAAAGCTCTA 2391  
 Db :|||  
 902 AspAspSer-----LeuArgLysGluLeuGluLysSerLysIle-----AsnLeu 916  
 Qy :|||  
 2392 ACCCCATGTTATGGACAA 2409  
 Db :|||  
 917 ThrAspAlaTyrSerGln 922

RESULT 6  
 B44450  
 N:ubiquitin-specific proteinase (EC 3.4.19.1) UBP3 - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YER151c; ubiquitin-specific processing proteinase  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
 C:Accession: B44450; S50654; S27467  
 R:Baker, R.T.; Tobias, J.W.; Varshavsky, A.

J. Biol. Chem. 267, 23364-23375, 1992  
 A:Title: Ubiquitin-specific proteases of *Saccharomyces cerevisiae*. Cloning of UBP2 and U  
 A:Reference number: A44450; MUID:93054674; PMID:1429680  
 A:Accession: B44450  
 A:Molecule type: DNA  
 A:Residues: 1-912 <BAK>  
 A:Cross-references: EMBL:M94917; MID:gl73129; PIDN:AAA35191.1; PID:gl73130  
 A:Note: sequence extracted from NCBI backbone (NCBIP:118310)  
 R:Dieterich, F.S.  
 Submitted to the EMBL Data Library, December 1994  
 A:Description: The sequence of *S. cerevisiae* cosmids 8229, 9115, 9132, 9981, and lambda  
 A:Reference number: S50654  
 A:Accession: S50654  
 A:Molecule type: DNA  
 A:Residues: 1-912 <DIE>  
 A:Cross-references: EMBL:U19917; MID:g603377; PID:g603391; MIPS:YER151c  
 C:Genetics:  
 A:Gene: SGD:UBP3  
 A:Cross-references: SGD:S0000953; MIPS:YER151c  
 A:Map position: 5R  
 C:Keywords: hydrolase

Alignment Scores:  
 Pred. No.: 0.00121 Length: 912  
 Score: 159.50 Matches: 154  
 Percent Similarity: 33.14% Conservative: 130  
 Best Local Similarity: 17.97% Mismatches: 324  
 Query Match: 2.07% Indels: 249  
 DB: 2 Gaps: 31

US-09-874-162A-7 (1-4409) x B44450 (1-912)

Qy 20 GCCACACACCGGGGGCTCTCGATGTAGCACCATGACAGCAGTCCGCGCGCTCTTCT 79  
 Db :|||||  
 6 AlaAsnLysGluGluSerTyrSerMetTyrProLysThrSerSerProProProPro 25  
 Qy 80 TCTCAATACCTGCCGANTCGGGGGTTCGGACTCCACTTCCCAACCTCCGCGCGCTCA 139  
 Db :|||||  
 26 -----ProThrAsn 28  
 Qy 140 TCGAGCACATCGAGGACCAACCATCGATACATGATCCACGCGTTTAGAAAACAAGAAT 199  
 Db :|||||  
 29 MetGlnIleProIleTyrGlnAlaProLeuGlnMetTyrGlyTyr--ThrGlnAlaPro 48  
 Qy 200 TACAGCAGCAACCACTATGTTGCCCTGAGTTACATAATAGATTATCATGACAGCTGCC 259  
 Db :|||||  
 48 YrLeuTyrProThrGlnIleProAlaTyrSerPheAsnMetValAsnGlnAsnGlnPro 68  
 Qy 260 GCCAGAGCAGGAGTCCCTAAAGAGAAGATTACGCGGAAGCTCTCGCTGACTCTGCTCA 319  
 Db :|||||  
 68 leTyrHisGlnSerGlySerProHisIleuProGlnAsnAsnIleAsnGlyGly 88  
 Qy 320 GCTCAGTGTCTCGAGGGAATGTCTCCACTCCCAACGCGCAGCAGTGGAGGCTTACTC 379  
 Db :|||||  
 88 erThrThrAsnAsnAsnIleAsnLysLysLysTrpHisSerAsnGlyIleThrAsnA 108  
 Qy 380 CCCCCGTGACCCCACTCCCTCTCTCTTTCATCTCCGACGACGACTCCGACAGAGC 439  
 Db :|||||  
 108 snAsnGlySerSerGlyAsnGlnGlyAlaAsnSerSerGlySerGlyMetSerTyrAsnL 128  
 Qy 440 CAACACAGATCTATAGATTCTTCGAACTCGGAATCTCATAGCACCATAATTTTGCACA 499  
 Db :|||||  
 128 ysSerHisThrTyrHisAsnTyrSerAsnAsnHisIle----- 141  
 Qy 500 GAACCTTTACTATGCTCTCATCGAAACTCCAGAAACAAACATCAAAAGGAAAAACATTTA 559  
 Db :|||||  
 142 -----ProMetMetAlaSerProAsnSerGlySerAsn----- 152  
 Qy 560 AGTTTCATGATATGTTTCAAAAGTAGAGAAAATGAAGAGAGAGAGAGAAATCTCATAGCT 619  
 Db :|||||  
 153 -----AlaGlyMetLysGlnThrAsnSerSerAsnGlyAsnGly 167  
 Qy 620 TGTGAGTCAATTTGCGAGCTTACGTTTCTTCTTCCACAAAATGATAGCCCATCAC 679

```

Db      167  erSerAla-----|||||-----ThrSerProSerT 174
Qy      680  CAAACTCAGAAATGAAACAAATTTCTGTACCTGGAGTCTCTGTGTGAAAGTTGGCC 739
Db      174  yrSerSerTyraSerSerSerGlnTyraSp-----LeuTyrllysPheAspV 190
Qy      740  ACAAAAAGAAAGAGATGATAGTTGTCATTAAGCGAAGTCCACACAGGTAAAGACAGG 799
Db      190  alThrLysLeuLysAsn-----LeuLysGluAsnSerSerAsnLeuLeuGlnL 206
Qy      800  TGCCCTTTGATTCCTGACCTCAATCAACAAACCCCGGA-----837
Db      206  euProLeuPhe-----lleAsnThrThrGluAlaGluPheAlaAlaSerValGlnA 224
Qy      838  -----AATTCCTGGTCCCTTGCGAGTTTCCAGTAAATGAATTTGAACCTAGTAAC- 885
Db      224  rgTyrGluLeuAsnMetLysAlaLeuAsnLeuAsnSerGluSerLeuGluAsnSerS 244
Qy      886  -----AGCCATATGCTGAAAGTCTTACTCGTTCGTATTTAGAGTGACTC 928
Db      244  alGluLysSerSerAlaHisHisThrLysSerHisSerIle-----258
Qy      929  GTCCAGGAAGAGAGAGATTTAATGAATGATTAATGGAGAAACCAAT---GAAATATTG 985
Db      259  -----ProLysHisAsnGluGluValLysThrGluThrHisGlyGluGluGluA 275
Qy      986  ATGTCATTAAGAGAGCTCCAGCCAGAGAAACGAAATCGTGAGAGTGGGAAAAGACAT 1045
Db      275  spAlaHisAspLysLysProHisAlaSerLysAspAlaHisGluLeuLysLysThr 295
Qy      1046  TTGTTGCAAAATGACAGTATTGTGATAAAACAGCGCTTTACAGCTTTTAGATGGGA 1105
Db      295  luValLysLys-----298
Qy      1106  ATGAAGTACCATGAGGAATGGAAGATGCTCAATTAAGCAAGAAAGACACATGGG 1165
Db      299  --GluAspAlaLysGlnAspArgAsnGluLysValIleGlnGluProGlnAlaThrVal 318
Qy      1166  AGACTATTCTTGATGGGAGAGGCTCGCTCCATTCGAAACATTTTCTCAGGACCTACGT 1225
Db      318  euProValValAspLysLysGlu-----ProGluGluSerValGluGluAsn-----333
Qy      1226  TCAGTTCACCTTCGTTGGACAGAGACCAATGATTAATCTACGGCTCTCTATTGCC- 1284
Db      334  -----ThrSerLysThrSerSerProSerProSerProSerProAlaA 347
Qy      1285  -----AAACCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACCAAGC 1333
Db      347  laLysSerTrpSerAlaIleAlaSerAspAlaIleLysSerArgGlnAlaSerAsnLys 367
Qy      1334  CTGGTTCAAGTAAACCTACTCAAACTATGCTGTTAAAGATCAATCTGACTACAGATCTAC 1393
Db      367  hrValSerGlySerMetValThrLysThrProIleSerGlyThrThrAlaGlyValSer 387
Qy      1394  AAACAGAAAAGAAAGGATCTCCAAATGAAACCGACAAAATTAAGATATTTTATC 1453
Db      387  erThrAsnMetAlaAlaThrIleGlyLysSerSerProLeu-----402
Qy      1454  AGTTTCTCTATAACAACTACAAGGCAACAACTAGAGCAAGACGATCACTGCTATGCC 1513
Db      403  -----LeuSerLysGlnProGlnLysLysAspLysLysLysValP 416
Qy      1514  CTGTGTACTCTGACTCCGCCCAACTTATAGTTTACTCAAGCATCTTAACTCTGCC 1573
Db      416  roProSerThrLysGlyIleGluProLeuGlySerIleAla-----LeuArgMetCys 434
Qy      1574  ATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGATGTTCTA 1633
Db      434  heAspProAspPheIleSerTyValLeuArgAsnLysAspValGluAsnLysIlePro 454
Qy      1634  TCAATGAGTGTATTGATGGCTCTCTATGACGAGGAATCCTCAGGATATTCATGCCAAC 1693

```

```

Db      454  alHisSerIleIle-----ProArgGlyIleIleAsnArgAlaA 467
Qy      1694  GATTTCCTTTTAGTCCCAACGACAGTAAAGAGAACACCTATCACATATTTCTGTGT 1753
Db      467  snIleCysPhe-----MetSerSerValLeuGlnValLeuLeuTyrc 481
Qy      1754  CGAGGCCA-----AAACGAAACAAAGCAAGCATG- 1782
Db      481  yslsYsPheIleAspValIleAsnValLeuSerThrArgAsnThrAsnSerArgValG 501
Qy      1782  -----1782
Db      501  lyThrSerSerCysLysLeuLeuAspAlaCysLeuThrMetTyrllysGlnPheAspLys 521
Qy      1783  -----TCTCAATTTCTTGAATCTGAAGATGGG-----GAACTAGAACACAGCAA 1825
Db      521  luThrTyrllysLysPheLeuGluAsnAlaAspAlaGluLysThrThrGluSerA 541
Qy      1826  GAACATATAGTAGTGGCCCAATCGTCTGTATTTCCNTAGTAGTACCTGCTTACCTCTCC 1885
Db      541  spAlaLysLysSerLysSerLysSerPheGlnHisCysAlaThrAlaAspAlaValL 561
Qy      1886  GTCCCAACAAATGGAAGTAGATAGTGAAGATGAAAGATCCTCAATGGCTAAGAGAAA 1945
Db      561  ysProAspGluPhe-----Tyrl 567
Qy      1946  AAACCATTAACAAATTGAAGATTTTCTGTATTTAATGAAGAGAGAGAAAGATGATGA 2005
Db      567  ysThrLeuSerThrIleProLysPheLysAspLeuGlnTrpGlyHisGlnGluAspAla 587
Qy      2006  AACTCTGGATCTCCATGTCATGAACCAT-----GGGT 2038
Db      587  luGluPheLeuThrHisLeuLeuAspGlnLeuHisGluLeuLeuSerAlaIleAspG 607
Qy      2039  TTATTCCTCAGCAATCAATGAATCATGCTGTATGCTGTTGTGAAATTTATGACAGA 2098
Db      607  lyleuThrAspAsnGlu-----IleGlnAsnMetLeuGlns 619
Qy      2099  AATAATTAAGAAAGATTTATGCTGAAACTTTCATCTTCATCTAGTCAGCATGCAT- 2154
Db      619  erIleAsnAspGluGlnLeu---LysValPhePheIleArgAsnLeuSerArgTyrl 638
Qy      2155  ----CACTTTAATCTTATTAGCATAATGTCATAGATAAAGCTGTTACCAAGCTCCGT 2209
Db      638  ysAlaGluPhe-----IleLysAsnAlaSerProArgLeuLysG 651
Qy      2210  AAATGACAGAAAATTTAGAAAAGGGGAATCTGCTCCCTCCCAACAGAGAAATACTG 2269
Db      651  luLeuIleGluLysTyrl-----GlyValIleAsnAspAspSerThrG 665
Qy      2270  AGAACAATAATGGACAGCAAAATGGATTTAGTGAATTAACCTCAAAAGAGAAAGCTTGG 2329
Db      665  luGlu-----AsnGlyTrpHisGlu-----671
Qy      2330  AAACAGATAGTGTCTCAGGGGTTTCAAAACAGAGCAAGCAAAACAAAAA 2376
Db      672  -----ValSerGlySerSerLysArgGlyLysLysThrLys 683

```

## RESULT 7

T15496  
 hypothetical protein C14F5.3 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 R;Accession: T15496  
 R;Minx: P.  
 submitted to the EMBL Data Library, June 1995  
 A;Description: The sequence of *C. elegans* cosmid C14F5.  
 A;Reference number: Z18361  
 A;Accession: T15496  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1263 <MIN>  
 A;Cross-references: EMBL:U29082; NID:G861384; PID:G861386; PIDN:AAA68402.1; CESP:C14F5.3









1765 --CGAACAAAGCA-----AG 1778

656 lHisSerIleLysProIleAspLeuGluAsnThrLysSerArgGlnGlnAlaIleLysAs 676

1779 CATGTCGTGAATTTCTGAATCTCAAGATGGGAAGTAGAACAAGCAAGAAACATATAGTAG 1838

676 pLeuAsnGluPheLeuLysAsnProAsnAspAlaGlnAlaSerLysThrLeuAlaG 696

1839 TGCCCACAAT-----CGTCTGTAATTCATAGTACTACTCGT 1877

696 nAlaAsnLysIleGlnHisLeuGluAspLeuLysSerLysValHisSer----- 712

1878 ACCTCTCCGTCCTCCACAGAAATGGAAGTAGTAGTAGAAGATCAAAAGCATCTGAATGGCT 1937

713 ---IleLysPro-----IleAspLeuGluAsnThrLysSer----- 723

1938 AAGAGAAAAACCATTACACAAATTCGAAGTTTTCTGATGTTAATCAAGAGAGAAAGA 1997

724 -ArgGlnGlnAlaIleLysAspLeuAsnGluPheLeuLysAsnProAsnAspAlaG 743

1998 AGTGATGAAACTCTGGAATCTCCATGTCATGAAGCATGGGTATTATCTGCACAAATCAAT 2057

743 nAlaSerLysThr-----LeuAlaGlnAlaTyrg 753

2058 GAATCATGCTGTATGCTGTTTGTAGAAAATTTGACAGACAAAATAATTAAGAAATTT 2117

753 uAsnAsnGlyAspLeuLeuLysAlaGluAsnAlaTyrgLysIleIleLysLeuThrAs 773

2118 ATTCGCA-----AACTTCATGCTT-----CA 2138

773 nThrGlnGluAspHisTyLysLeuGlyIleIleArgPheLysLeuLysTyrgLys 793

2139 TCTAGTCAGATGCATGCATTTAATCTT----- 2166

793 sSerIleGluSerPheAspGlnThrIleLysLeuAspProLysHisLysLysAlaLeuHi 813

2167 -----ATTAGCATATGTCATAGATAAGCTGTTACCAAGCTCCGTGAAATGCA 2216

813 sAsnLysGlyIleAlaLeuMetMetLeuAsnLysAsnLysLysAlaIleGluSerPheG 833

2217 GCAA-----AAATTAGAAAAGGGGAATCTGCTCCCTGCCAAACGAGAAATAACTGA 2270

833 uLysAlaIleGlnIleAspLysAsnTyrgLysThrAlaTyrgLysGlyIleAlaG 853

2271 AGAACAAAATGGACAGCAATGATTTAGTGAATTAACCAAAAGAGAAAGCTTTGGA 2330

853 uGlnLysAsnGly----- 857

2331 AACAGATAGTGTCTCAGGGGTTTCAACACAGACAGAAAAACCAAAACTCTGAAAAGCTCT 2390

858 ---Asp-MetGlnGlnAlaPhe-----AlaSerPheLysAsnAlaTyrgAsnLeu- 872

2391 AACCCCATGTTATGGCAACACATGAAATTAATT----- 2425

873 -----AspLysAsnProAsnTyrgAlaLeuLysAlaGlyIleValSerAsnA 888

2426 --TTAGGGAAT---TCATCCTCTAAGAAATATGTTTTGTTTTTAAT 2467

888 snLeuGlyAsnPheLysGlnSerGluTyrgLysLeuAsnPhePheAsn 903

RESULT 9

T39068

coiled coil protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T39068

R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Connor, R.E.; Wood, J.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z21825

A:Accession: T39068

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

139068 coiled coil protein - fission yeast (Schizosaccharomyces pombe)

C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 03-Dec-1999

R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Connor, R.E.; Wood, V.

A;Accession: T39068

A;Molecule type: DNA

11/10/2000 11:50:00 AM

A:Residues: 1-1208 <MUR>  
 A:Cross-references: EMBL:Z81317; PTDN:CA803608.1; GSPDB:GN00066; SPDB:SPAC6G9.06c  
 A:Experimental source: strain 972h-; cosmid c6G9  
 C:Genetics:  
 A:Gene: SPDB:SPAC6G9.06c  
 A:Map position: 1

Alignment Scores:  
 Pred. No.: 0.00192 Length: 1208  
 Score: 157.00 Matches: 184  
 Percent Similarity: 35.51% Conservative: 153  
 Best Local Similarity: 19.39% Mismatches: 300  
 Query Match: 2.04% Indels: 312  
 DB: 2 Gaps: 40

US-09-874-162A-7 (1-4409) x T39068 (1-1208)

QY	127	CTGGCCGACCTCATCGACGACGAGTCCCTAAAGAGAGATTTCAGCCGAGCTCTCG	186
Db	127	LeuThrAspLeuSerLysAsnSerGluIleAspHisThrAspGly	141
QY	187	GAATAACAAGATTACAGCAGCCACCTATGTTCCCTGAGTTACATAAATAGATTCTG	246
Db	142	-----GluLeuProAlaAsnAlaLeuThrLeu	151
QY	247	ACAGATGTCGCCCGCAGAGCAGGAGTCCCTAAAGAGAGATTTCAGCCGAGCTCTCG	306
Db	152	-----ArgGluGlnGluLysValLeuGluLysVal-SerArgGluAsnPh	166
QY	307	CTGACTC---TGTCAGCTCAGTGTCTCGAGGGAATGTCTCACTCCCGCAGCCAGC	363
Db	166	eGlyLeuArgIleLysIleValCysLeuGlu-LysArgLeuGluSerMetAlaProGluG	186
QY	364	AGTGGAGCCTTACTCCCGCCGTCAGCCGCCACCCATCCCTCTCTTATCCGCGAGC	423
Db	186	inileLysGlu---AlaValLysAspAsnVal	197
QY	424	AGCCTCCGACGACGACCAACACGATCTATGATTTCTCGAACTCGGAATCTCATAGCA	483
Db	197	euHisAlaGluArgAlaAsnLeuGlnLeu-GlnLeuLysArgThrGluSerLeuLeu---	215
QY	484	CAATATTTTTCACAGAACTCTTACTTACATGTCTCATCGAACTCCAGNACAAACATC	543
Db	216	-----GlnLysSer	218
QY	544	AAAAGGAAACATTTAAAGTTGAT-----GATATGTTATCAAAAGTAGAGAAAATG	594
Db	219	GluAspLysAsnPhLeuGluGluLysValAspTyrLeuSerLysValAsnAspVal	238
QY	595	AAAGGAGAGCAAGATCTCATAGCTGTTCAGCTCATTTGTCAGCTTACGTTTACTGGTTTC	654
Db	239	GluGlnSerGlnAsnValLysValPheThrGluArgIleArg-----	252
QY	655	TTCCCAAAAAT-----GATAAGCCATCACCACAACTCAGAAAAT	693
Db	253	PhLeuGluAsnAlaLeuGluLysValGlnArgGluLysAspSerLeuSerThrGluMet	272
QY	694	GAACAAAATTCGTTTACCTCGAGTCTCTGCTGTGTGAAAGTTTCCACAAAAAGAAAG	753
Db	273	GluGluAspLysSerAsnLysGluVal-----	281
QY	754	GATGTAGTTGTCCCAATAGCAGTGTCCACAGGT-----AAAAACGAGTG	801
Db	282	AspTyrGluTyrGluIleArgGlnLeuGlnAsnArgLeuAspGluLeuSerGluGluLeu	301
QY	802	CCTTGATTCTCTGACCTCAATCAAAACAAACCGGAAATTTCCCGTCCCTTCGATTTC	861
Db	302	AspValAlaGlnAspLeuLeuThrGluLysGluAspGluIleAlaThrLeuLysArgGln	321
QY	862	AGTAATGTAATTGACCTAGTAACAGC-----	888
Db	322	IleGluGluLysGluAsnSerSerSerAlaPheGluAsnGluGluAsnSerSerTyrVal	341

QY	889	CATATGGTGAAGTCTTACTCTGTTCTGATTTAGAGTGTGCTGCCAGGAAGAAGAGAGTTT	948
Db	342	HisLeuGlnGluAspTyrAlaIleLeu-----	350
QY	949	AATGAATGATTATATGAGAACCAATGAAATATTTGATGTCATATGATGAAGACTTCCAGCC	1008
Db	351	-----GlnAlaLysCysAspGluPheAlaAspArgIleGlnValLeuThrAla	366
QY	1009	AGAAGAAAACGAAATCGTGAG-----	1029
Db	367	AspLeuGluLysGluLysGluAsnGlnIleMetHisGluSerGluAlaSerIleGlyLeu	386
QY	1030	---GATGGGAAAAGACATTTGTCACAAATGACAGTATTGTATATAAAACAGCGCTTA	1086
Db	387	ThrAspSerMetGlnValHisThrLeuGlnGluLeuHisLysAlaAsnGluGluLeu	406
QY	1087	CAGCTTTTATGATGGGAATAT-----GAAGTA	1113
Db	407	GluPheLeuHisAspGlnIleSerArgMetAsnGluGluLysAsnPhGluAspIle	426
QY	1114	GCCATGCGAG-----GAAATGGAAGAAATGTCACAAATGACCAAGAAAGAGCAACA	1161
Db	427	MetLeuGlnPheArgSerLeuGluGluGluArgAspValLeuGluSerLysLeuGlnThr	446
QY	1162	TGGGAGACTATTCTTGATGGGAGAGGCTGCTCATTCGAAACATTTTCTCAGGACCT	1221
Db	447	LeuGluAspAspAsnSerLeuArgLeuMetThrSerSerLeuGlyAsnGlnIleGlu	466
QY	1222	AGTTGTCAGTTCCTCTTCGTTGGACAGAGAG-----	1254
Db	467	SerLeuArgThrGlnAsnArgGluLeuAspGluGluLysAsnHisLeuArgLeuLeuAla	486
QY	1255	-----ACCAATGATAAATCTACGGCT-----CCTATTGCCAAA	1287
Db	487	SerLysAsnSerAspLysAlaLeuAlaGluThrAsnIleArgLeuGlnGluValThrLys	506
QY	1288	CCTCTGCGACT-----AGAAATTCAGAGAGTCTCCAT-----	1320
Db	507	GluLeuGluThrLeuArgMetLysAsnSerAsnAspLeuAsnGluIleHisAspLeuArg	526
QY	1321	CAGGAAAACAAAGCCTGTTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCAT	1380
Db	527	GluGluAsnGluGlyLeuThrLeuLysIleAspSerIleThrLysGluLysAspArgLeu	546
QY	1381	ACTACAGATCTCAAAACAAGA-----AAAGAAAGGAT	1413
Db	547	IleAsnGluLeuGluGlnArgIleLysSerTyrGluValAsnValSerGluLeuAsnGly	566
QY	1414	ACTCAAAATGAAACCGACAAAATTAAGA-----ATATTT	1449
Db	567	ThrIleAspGluLysArgAsnLysLeuLysAspLysGluGluThrTyrAsnGluValMet	586
QY	1450	TATCAGTTTCTCTAT---AACCAACAATACAGCAACAACTGAAAGCAAGAGATGACCTG	1506
Db	587	AsnAlaPheGlnTyrLysAspAsnAspLeuArgPheHisGluSerIleAsnLysLeu	606
QY	1507	CATTGCCCTTGTGTACTCTGACTGCCGCAACCTTATAGTTTACTCAAGCATCTTAAA	1566
Db	607	GlnAsp-----ArgGluLysGluLeuThrSerAsnLeuGluLys-LysAs	621
QY	1567	CTCTGCGCATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAAGGTCTAGTAGAT	1626
Db	621	nLeuValIleSerSerLeuArgThrValAlaMetLeuGluLys-----	636
QY	1627	GTTTCTTCAATGAGTGTATGATGGCTCTATGAGGAAATCTCTCAGGATTTTCATCGC	1686
Db	637	-----GluArgGluSerIleLysTyrLeuSe	646
QY	1687	C-----AACCTGATTGCTTTTGTAGTCGCAACCGACCCAG-----	1720
Db	646	rGlyAsnAlaLysAspLeuAspAsnThrAsnLeuMetGluIleLeuAsnAspLysIleSe	666

Db 666 rValLeuGlnArgGlnLeuThrAspValLysAspGluLeuAspValSerGluGluGluAr 686  
 QY 1745 -----TTCATTGTGTGCAGGCGCAAAACGCAAAAGCAAGCATGCTGCAATTTCTTGA 1797  
 Db 686 gGluGluAlaIleValAlaGlyGln--LysLeuSerAlaSerPhe--GluLeuMetSer 704  
 QY 1798 TCTGAAGATGGGGAAGTGAACAGACAGCAATATAGTATAGTGGCCCAATCGCTGTGAT 1857  
 Db 705 AsnGluLysGlnAlaLeuGluLysLys-----TyrSerSerLeuLysAsnGluLeu--- 721  
 QY 1858 TTCATAGTATACCTGCTTACCTCCGTCACAGAAAGTGAAGTAGATAGTGAAGAT 1917  
 Db 722 IleAsnAlaGlnAsnLeuLeuAspArgGluGluGluLeuSerGluLeuSerLysLys 741  
 QY 1918 GAAAGGATCTGGAATGAGTGAAGAAACCAATACACAAATGGAAGTCTTCTGAT 1977  
 Db 742 LeuPheGluGluArgLysIleArgSerGlySerAsnAspAspIleGluLysAsnLysGlu 761  
 QY 1978 GTTATGAGGAGAGAGAGATGATG--AACTCTCGAATCTC--CATGTCTAGNAG 2031  
 Db 762 IleAsnValLeuAsnSerGluLeuAlaAspLysLeuAlaGlnIleArgHisLeuGluSer 781  
 QY 2032 CATGGTATTATGCTGACAATCAATGAATCATGCTGCTGCTGTTGTAGAA---AAT 2088  
 Db 782 AspLysMetGluLeuAspLysLeuValHisHisLeuAsnArgGlyIleGluGluAlaAsn 801  
 QY 2089 TATGACACAAAATTAATGAAGATTTATGTCGAAATCTCATCTCTATCTAGTACG 2148  
 Db 802 IleGluGluAsnAlaValLysLysArgLeuCys-----LeuLeuMet 815  
 QY 2149 ATGCATGACTTTAATCTTATAGCATAATGTCATA---GCTGCTGCTGCTGCTGCTGCT 2184  
 Db 816 GlyCysAspTyrSerSerValSerIleLeuGlnIleValSerGlnIleGluHisPheVal 835  
 QY 2185 GATAAGCTGTTACCAAGCTCCGTCGAAATGCGACGCAAAATAGAA--- 2229  
 Db 836 AsnGlnGlnIleGlnThrIleArgSerLysGlnGluLeuArgHisAspPheValGln 855  
 QY 2229 ----- 2229  
 Db 856 PheSerGlyLysLysGluGlnGluLeuSerArgSerPheGluLysPheGlyLeuGlyThr 875  
 QY 2230 -----AAGGGGGAATCTGCTTCCCTCGAACGAAAGAAATCACTGAAGAACAAAT--- 2280  
 Db 876 GluThrLysHisAspIleLeuAlaGlnArgAsnArgAsnValSerGluLysMetAsnAsp 895  
 QY 2281 -----GGGACAGCAAAATGGATT--- 2288  
 Db 896 LeuGluAsnAlaAlaGlnLysPheSerSerProAspArgLysAsnGlyTyrLeuTyr 915  
 QY 2299 -----AGTGAATTTAACTCAAAAGAGAAAGCTTTGCAA--- 2331  
 Db 916 ProSerGluHisThrSerLysIleGluTyrLeuGluLysThrIleGluAspLeuLysLeu 935  
 QY 2332 -----ACAGATAGTGTCTCAGGGGTTTCA 2355  
 Db 936 AlaLeuGlnAspGluLeuLysAsnArgAsnLeuMetAspAspIleSerSerTyrAsn 955  
 QY 2356 AAACAGACCAAAAA 2370  
 Db 956 LysGlnThrThrLys 960

RESULT 10  
 S69838  
 TVB protein - yeast (Saccharomyces cerevisiae) retrotransposon Tyl.GRI  
 N/Alternate names: protein G4054  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 09-Mar-1996 #sequence\_revision 06-Sep-1996 #text\_change 20-Jun-2000  
 C/Accession: S69838  
 R/Rieger, M.; Mueller-Auer, S.; Bruckner, M.; Schaefer, M.  
 submitted to the Protein Sequence Database, May 1996  
 A/Reference number: S64071

A/Accession: S69838  
 A/Molecule type: DNA  
 A/Residues: 1-1755 <RMB>  
 A/Cross-references: EMBL:Z78212; NID:G1325961; PIDN:CAA97012.1; PID:G1325964  
 A/Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4  
 C/Genetics:  
 A/Map position: 7R  
 A/Mobile element: retrotransposon Tyl.GRI  
 C/Superfamily: TYB protein

Alignment Scores:  
 Pred. No.: 0.00239 Length: 1755  
 Score: 156.00 Matches: 184  
 Percent Similarity: 35.44% Conservative: 124  
 Best Local Similarity: 21.17% Mismatches: 283  
 Query Match: 2.03% Indels: 278  
 DB: 2 Gaps: 48

US-09-874-162A-7 (1-4409) x S69838 (1-1755)  
 QY 163 ATGATACAGATCCAGGGTTTAA-----GAAAAACAGAATTACAGCGCCCAACC 213  
 Db 881 ValAspThrThrAsnTyrValIleLeuGlnGlyLysGluSerArgLeuAspGlnPheAsn 900  
 QY 214 TATGTTGCCCTGAGTTAC-----ATAAATAGATTTCATGACAGAT----- 252  
 Db 901 TyrAspAlaLeuThrPheAspGluAspLeuAsnArgLeuThrAlaSerTyrHisSerPhe 920  
 QY 253 ---GCTGCCCGCAGAGCAGGAGTCCCTAAAGAGAAGATT----- 291  
 Db 921 IleAlaSerAsnGluIleGlnGluSerAsnAspLeuAsnIleGluSerAspHisAspPhe 940  
 QY 292 -----CAGCCGAAGCTCTCGCTACTGCTCCAGTCA 324  
 Db 941 GlnSerAspIleGluLeuHisProGluGlnProArg-----AsnValLeuSerLysAla 958  
 QY 325 GTGCTCGAGGAATGTGTCACCTCCCCCA-----CGCCAC 360  
 Db 959 ValSerProThrAsp---SerThrProProSerThrHisThrGluAspSerLysArgVal 977  
 QY 361 AGCAGTGAAGCCTTACTCCCCC-----GTGACCCGCCCATCACCCTCC----- 408  
 Db 978 SerLysThrAsnIleArgAlaProArgGluValAspProAsnIleSerGluSerAsnIle 997  
 QY 409 -----TCTTCATTCGCGACGACGACTCCGACAGACCAACACAGATCTATAGATTCTT 462  
 Db 998 LeuProSerLysLysArgSerSerThrPro-----GlnIleSerAsnIleGlu 1013  
 QY 463 CGAATCTCGGAAT-----CTCATAGCACCAATATTTTTC 495  
 Db 1014 SerThrGlySerGlyMetHisLysLeuAsnValProLeuLeuAlaProMetSerGln 1033  
 QY 496 CACAGACTCTTACTTACATGCTCTCGAACTCCGAGAACATCAACATCAAGGAAACA 555  
 Db 1034 SerAsnThr-----HisGluSerSerHisAlaSer---LysSerLysAsp 1047  
 QY 556 TTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAAATGAAAGGAGCAAGAAATCTCAT 615  
 Db 1048 PheArgHisSerAspSerTyrSer-----GluAsnGlu 1058  
 QY 616 AGCTTGTCAGCTCATTTGCGAGCTTACTGTTTCTTCTCCACAAAATGATAAGCCA 675  
 Db 1059 ThrAsnHisThrAsnValProIleSerSerThrGly-----GlyThrAsnLysThr 1076  
 QY 676 TCACCAACTCAGAAAATGAACAAATCTGTTACCTCGAAGTCTGCTTGTGAAAGTT 735  
 Db 1077 ValProGlnIleSerAspGlnGluThrGlu-----LysArgIle 1089  
 QY 736 TGGCACAATAAAGAAAG---GATGTAAGT----- 762  
 Db 1090 IleHisArgSerProSerIleAspAlaSerProProGluAsnAsnSerSerHisAsnIle 1109  
 QY 763 TGTCATAAGGCAAGTTCCACAGGTAAACACAG----- 798

```

Db      1110 ValProIleLysThrProThrThrValSerGluGlnAsnThrGluGluSerIleIleAla 1129
QY      799 ---GTGCTTTGATTCCTGACCTCAATCAAAACAAACCCGAAATTTCCCGTCCCTTGCA 855
Db      1130 AspLeuProLeu---ProAspLeuProProGluSerProThrGluPheProAsp----- 1146
QY      856 GTTCCAGATATGAATTTGAACCTAGTAACAGCCATATGGTGAAGTCTTACTCGTCTCTA 915
Db      1147 ---ProPheLysGluLeuProProIleAsnSerArgGlnThrAsnSer----- 1161
QY      916 TTTAGAGTGACTCGTCGACGAGAGAGAGAGTTAATGGAATGATTAATGGAGAACCAAT 975
Db      1162 -----SerLeuGlyGlyIle-----GlyAspSerAsn 1170
QY      976 GAAATATTGTGATCAATCAAGAGCTTCCAGCCAGAGAAACGAAATCGTAGGATGGG 1035
Db      1171 AlaTyThrThrIleAsnSer-----LysLysArgSerLeuGluAspAsn 1185
QY      1036 GAAAGACATTT---GTTGCACAAATGACAGATTTTGTATATAAAGACGGCGCTTACAGCTT 1092
Db      1186 GluThrGluIleLysValSerArgAspThrTrpAsnThrLysAsnMetArg----- 1202
QY      1093 TTAGATGGGGAATATGAAGTAGCATGAGGAGGAGGAAATGGAAGATGTCCTCAATTAAGCAAGAA 1152
Db      1203 -----SerLeuGluProProArgSerLysLys 1211
QY      1153 AGA-----GCAACATGGGAGACTATTTGTATGGAGAGAGGCTGCCCTCCATTC 1200
Db      1212 ArgIleHisLeuIleAlaValIleVal-----LysSerIleLysProIle 1228
QY      1201 GAACATTTCTCAGGACCTAGCTGTCAGTTC-----ACTCTTCGTTGGACAGGAGAG 1254
Db      1229 ArgThr-----ThrLeuArgTyThrAspGluAlaIleThrTyAsnLysAsp 1243
QY      1255 ACCAATGATAAATCTACGGCTCTTATTGCCAAACCTCTTGCCACTAGAAATTCAGAGAGT 1314
Db      1244 IleLysGluLysGlu-----LysTyIleGluAla 1253
QY      1315 CTCCTCAGGAAACACACCTGGTTCAGTTAACTACTCAAACTATGCTGTTAAAGAA 1374
Db      1254 TyHisLysGlu-----ValAsnGlnLeuLeuLysMetLysThr 1266
QY      1375 TCATTGACTACAGATCTACAAACAGAAAGAAAGGATACTCCAAATGAAACCGACAA 1434
Db      1267 TrpAspThrAspGluTyThrAspArgLysGluIleAsp-----Pro 1280
QY      1435 AAATTAAGAATATTTATCATGTTTCTCTATTAACAACATCAAGGCAACAACTGAAGCA 1494
Db      1281 LysArgValIleAsnSerMetPheIlePheAsn-----LysLys 1293
QY      1495 AGAGATGACCTGCATGCGCTTGGTGTACTCTGAACCTGCCGCAAACTTTATAGTTTACT 1554
Db      1294 ArgAspGlyThrHis----- 1298
QY      1555 AAGCATCTTAAACTCTGCCATPAGCAGATTTATCTTCAACTATGTTTATCTCCAAAGGT 1614
Db      1299 -----LysAlaArgPheVal----- 1303
QY      1615 GCTAGGATAGATGTTTCTCATCAATGAGTGTATGATGCTTCCTATGCTCAGGAAATCCTCAG 1674
Db      1304 AlaArgGlyAspIleGlnHisProAspThrTyThrAspSerGlyMetGlnSerAsnThrVal 1323
QY      1675 GATATTCATCCCAACCTCGGATTTGCTTTTAGTCGCAACGACCGACAG-----TT 1722
Db      1324 HisHisTyThrAlaLeuMetThrSerLeuSerLeuAlaLeuAspAsnAsnTyThrIleTh 1343
QY      1723 AAGAGAACACCTATCACACATTTCTGTGTGTCAGGCGCAACGAAACAAAGAACGAGCATG 1782
Db      1343 rGlnLeuAspIleSerAlaTyLeuTyLeuTyLeuAlaAspIleLysGlu----- 1359
QY      1783 TCTGAATTTCTTGAATCTGAAGATGGGGAAGTAGAAGACAGCAACATATAGTAGTGGC 1842

```

```

Db      1359 ----- 1359
QY      1843 CACAATCGTCTGTATTTCATATGATACCTGCTTACCTCTCCGTCACAGAAATGGAA 1902
Db      1360 -----LeuTyIleArgProProHisLeuGlyMetAsnAspLys----- 1373
QY      1903 GTAGATAGTGAAGATGAAGAGTCTGAATGGCTAAGAGAAACCAATACAC----- 1957
Db      1374 -----LeuIleArgLeuLysSerLeuTyrgLeuLysGlnSerGlyAlaAsnTrpTy 1392
QY      1958 ----AAATTGAAGAGTTTCTGATGTTAATGAGGAGAGAGAAAGTATGAAACTCTGG 2013
Db      1392 rGluThrIleLysSerTyLeu--IleGlnGlnCysGlyMetGluGluValArgGlyTrp 1411
QY      2014 AATCTCCATGTCATGAAGCATGGTTTATTGTGCAATCAATGAATCATGCCCTGTATG 2073
Db      1412 SerCys-----ValPheLysAsnSerGlnValThrIleCys--- 1423
QY      2074 CTGTTTGTAGAAATATTGACAGAGAAATAATTAAGAGAAATTTATGCGAACTTCATG 2133
Db      1424 LeuPheValAspAsp-----MetValLeuPheSerLysAsnLeuAsnSerAsnLysArg 1441
QY      2134 CTT-----CATCTAGTCAGCATGACATGACITTAATCTTATTAGCATATGTCAATAGAT 2187
Db      1442 IleIleAspLysLeuLysMetGlnTyThrAspThrLysIleIleAsnLeuGlyGluSerAsp 1461
QY      2188 AAAGCTGTTTACC---AAGCTCGTGAAATGCGACAAATAATTAGAAAAGGGGAATCTGCT 2244
Db      1462 GluGluIleGlnTyThrAspIleLeuGlyLeuGluIleLysTyThrGlnArgGlyLysTyMet 1481
QY      2245 TCCCTCGCAACAGCAAAATAACTCAAGAACAAATAATGGACAGCAAAATGGATTTAGTCAA 2304
Db      1482 LysLeuGlyMetGluAsnSerLeuThrGluLysIleProLysLeuAsn-----ValPro 1499
QY      2305 ATTAATCTCAAAAGAGAAAGCTTTGAAACAGATAGTGTCTCAGGGGTTTCAAAACAGAGC 2364
Db      1500 LeuAsnProLysGlyArgLysLeuSerAlaProGlyGlnProGlyLeuTyIleAspGln 1519
QY      2365 AAAAACAAAACCTCTGAA-----AAGCTCTAACCCCATGTTATCGCAACAA 2412
Db      1520 GlnGluLeuGluLeu-GluGluAspAspTyThrLysMetLysValHisGluMetGlnLys-- 1538
QY      2413 CTGAAATTTACATTTTAGGGAATTCCTCTCAAGAAATTTATGTTTGTTTTAAATCATAT 2472
Db      1539 -----LeuIleGlyLeuAlaSerTyValGlyTyThrLysPheArgPheAspLeuLe 1555
QY      2473 GTTCCAAACAGGCATGTT 2491
Db      1555 TyThrTyIleAsnThrLeu 1561

RESULT 11
S57045
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N;Alternate names: protein J1560; protein YJ027W
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
C;Accession: S57045; S60511
R;Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herber
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57040
A;Accession: S57045
A;Molecule type: DNA
A;Residues: 1-1755 <HUA>
A;Cross-references: EMBL:Z49526
A;Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residue 4;
R;Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herber
Yeast 11, 1179-1186, 1995
A;Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading fra
A;Reference number: S60503; MUID:96109930; PMID:8619316
A;Accession: S60511
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 'TISTTFT', 435-1755 <ZAG>

```

A: Cross-references: EMBL:X87297  
 A: Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995  
 A: Note: the difference at the amino end is due to translation from an incorrect reading  
 C: Genetics:  
 A: Map position: 10R  
 A: Mobile element: retrotransposon Ty1.UR  
 C: Superfamily: TyB protein

Alignment Scores:  
 Pred. No.: 0.0026 Length: 1755  
 Score: 155.50 Matches: 159  
 Percent Similarity: 36.70% Conservative: 99  
 Best Local Similarity: 22.62% Mismatches: 232  
 Query Match: 2.02% Indels: 213  
 DB: 2 Gaps: 44

US-09-874-162A-7 (1-4409) x 857045 (1-1755)

```

QY 163 ATCCATACAGATCCACGGGTTTAA-----GAAACAAAGAAATTACAGACGACCAACC 213
Db 881 ValAspThr-ThrAsnTyrValIleLeuGlnGlyLysGluSerArgLeuAspGlnPheAsn 900
QY 214 TATGTGCTCCCTGAGTTAC-----ATAATAGATTTCATGACAGAT----- 252
Db 901 TyrAspAlaLeuThrPheAspGluAspLeuAsnArgLeuThrAlaSerTyrGlnSerPhe 920
QY 253 ---GCTGCCCGCGAGAGCAGAGTCCCTAAAGAAAGAGAT----- 291
Db 921 IleAlaSerAsnGluIleGlnGluSerAsnAspLeuAsnIleGluSerAspHisAspPhe 940
QY 292 -----CAGCGAAGCTCTCCCTGACTCTGTCAGCTCTGTCAGCTCA 324
Db 941 GlnSerAspIleGluLeuHisProGluGlnProArg-----AsnValIleSerLysAla 958
QY 325 GTGTCTCGAGGAATGTGCCATCCCA-----CGCCAC 360
Db 959 ValSerProThrAsp---SerThrProProSerThrHisThrGluAspSerLysArgVal 977
QY 361 AGCAGTGGAGGCTTACTCCCGCC-----GTGACCCCGACCCCATCCCGCTCC----- 408
Db 978 SerLysThrAsnIleArgAlaProArgGluValAspProAsnIleSerGluSerAsnIle 997
QY 409 -----TCTTCATCCGAGCAGCAGCTCCGAGAGCCACACACATCTATAGATTTCT 462
Db 998 LeuProSerLysLysArgSerThrPro-----GlnIleSerAsnIleGlu 1013
QY 463 CGAATCCGGAAT-----CTCATAGCACCATAATTTTG 495
Db 1014 SerThrGlySerGlyGlyMetHisLysLeuAsnValProLeuLeuAlaProMetSerGln 1033
QY 496 CACAGACTCTTACTTACATGCTCATCGAAATCTCAGAACAAACATCAAAAGGAAACA 555
Db 1034 SerAsnThr-----HisGluSerSerHisAlaSer---LysSerLysAsp 1047
QY 556 TTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAAGGAGACGACGAATCTCAT 615
Db 1048 PheArgHisSerAspSerTyrSer-----GluAsnGlu 1058
QY 616 AGCTGTGTCAGCTTTCAGCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCA 675
Db 1059 ThrAsnHisThrAsnValProIleSerSerThrGly-----GlyThrAsnAsnLysThr 1076
QY 676 TCACCAACTCAGAAATGAAACAAATCTCTGTTACCTGGAAGTCTGCTGTTGAAAGTT 735
Db 1077 ValProGlnIleSerAspGlnGluThrGlu-----LysArgIle 1089
QY 736 TGCCACAAAAAAGAAAG---GATGTAAGT----- 762
Db 1090 IleHisArgSerProSerIleAspAlaSerProProGluAsnAsnSerSerHisAsnIle 1109
QY 763 TGTCCAAATAGGCAAGTCTCCACAGGTAAAGAGAG----- 798
Db 1110 ValProIleLysProThrThrValSerGluGlnAsnThrGluGluSerIleIleAla 1129

```

```

QY 799 ---GTGCTTTGATTCTCTGACCTCAATCAACAAACCCGAAATTTCCGTCCTCTGCA 855
Db 1130 AspLeuProLeu---ProAspLeuProGluSerProThrGluPheProAsp----- 1146
QY 856 GTTTCCAGTAATGAATTTGAACCTAGTACAGCCATATGCTGAAGTCTTACTCTGCTA 915
Db 1147 ---ProPheLysGluLeuProProIleAsnSerHisGlnThrAsnSer----- 1161
QY 916 TTAGAGTGAATCGTCCAGGAAGAGAGATTTAATGAATGATTATGAGAAACCAAT 975
Db 1162 -----SerLeuGlyIle-----GlyAspSerAsn 1170
QY 976 GAAATATTGATGTCATGAAGAGCTTCCAGCCAGAGAAACGAAATCGTGAGGATGGG 1035
Db 1171 AlaTyrThrThrIleAsnSer-----LysLysArgSerLeuLeuAspAsn 1185
QY 1036 GAAAGACATTT---GTTGCACAAATCACAGTATTTGATATAAACACAGCGCTTACAGCTT 1092
Db 1186 GluThrGluIleLysValSerArgAspThrTyrAsnThrLysAsnMetArg----- 1202
QY 1093 TTAGATGGGAATATGAGTAGCATCCAGGAATGCAAGAAATGCAATTAAGCAAGAAA 1152
Db 1203 -----SerLeuGluProProArgSerLysLys 1211
QY 1153 AGA-----GCAACATGGGAGACTATTCTTGATGGGAGAGGCTGCCTCCATTC 1200
Db 1212 ArgIleHisLeuIleAlaValLysAlaVal-----LysSerIleLysProIle 1228
QY 1201 GAAACATTTTCTCAGGACCTAGTTCAGTTTC-----ACTCTTCGTTCGACGAGAG 1254
Db 1229 ArgThr-----ThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAsp 1243
QY 1255 ACCAATGATAAATCTACGGCTCCTATTGCCAAACCTCTGCCACTAGAAATTCAGAGAGT 1314
Db 1244 IleLysGluLysGlu-----LysTyrIleGluAla 1253
QY 1315 CTCCATCAGGAAACAAAGCCTGTTTAAACCTACTCAAACTATTGTTGTTAAAGAA 1374
Db 1254 TyrHisLysGlu-----ValAsnGlnLeuLeuLysMetLysThr 1266
QY 1375 TCATTGTACTAGATCTACAAACAGAAAGAAAGATCTCCAAATGAAACCGACAA 1434
Db 1267 TrpAspThrAspGluTyrTyrAspArgLysGluLeuAsp-----Pro 1280
QY 1435 AAATTAAAGATATTTTATCAGTTTCTCTATAACAACAATACA-----AGGCAA 1482
Db 1281 LysArgValIleAsnSerMetPheIlePheAsnLysLysArgAspGlyThrHisLysAla 1300
QY 1483 CAAACTGAGCAGAGATGACCTGCATTCGCT-----TGTGTACTCTG 1527
Db 1301 ArgPheValAlaArgGlyAspIleGlnHisProAspThrTyrAspThrGlyMetGlnSer 1320
QY 1528 AACTGCCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCATAGCAGATTTATC 1587
Db 1321 AsnThrValHisHisTyrAlaLeuMetThrSerLeuSerLeuAlaLeuAspAsn----- 1338
QY 1588 TTCAACTATGTTTATCATCCAAAGGTCGTAGGATAGATGTTTCTATCAATAGTGTAT 1647
Db 1339 ---AsnTyrTyrIle-----ThrGlnLeuAspIleSer-----SerAlaTyr 1351
QY 1648 GATGCTCTCATCGAGAAATCTCAGGATATTCATCGCAATTCATCGCACTGGATTTGCTTTAGT 1707
Db 1352 -----LeuTyrAlaAspIleLysGluLeuLeuIleArgProProProHisLeuGly 1369
QY 1708 CGCAACGAGCAGTTAAGAGAACACCTTCATCACATATTTCTGTGTGAGCGCCAAAACGA 1767
Db 1370 MetAsnAspLysLeuIleArgLeuLysLysSerHisTyrGlyLeu-----LysGln 1386
QY 1768 ACAAGACAGACATGCTCTGAATTTCTTGAATCT-----GAAGATGGGAGTAGACA 1818
Db 1387 SerGlyAlaAsnTyrTyrGluThrIleLysSerTyrLeuIleLysGlnCysGlyMetGlu 1406

```



Db 645 -AsnAsnLysThrLysGluIleLeuLysSerLysAsnTyrLeuGlu----- 659  
 QY 1524 TCTGAACCCGCAAACTTTATAGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATT 1583  
 Db 660 -AsnGluLysArg-----ThrLeuGluGluLeuLysLeuArgGlyLysAsnAs 675  
 QY 1584 TATCTTC-----AACTGTTTATCATCTCAAAAGGTCTAGGATA----- 1623  
 Db 675 nilePheLysLysAspGluLysTyrAsnSerLeuGluValIleIleAsnGluIleG1 695  
 QY 1624 -----GATGTTCTCATCAATGAGTGTATGATGCTCTCTATGCAGGA----- 1665  
 Db 695 nileAsnGluGluAsnLysIleAsnAspIleGlnAspGlyAsnIleSerLysGlnLysI1 715  
 QY 1666 -----AATCTCAGGATATTCATCGCAACC 1691  
 Db 715 eileGlnSerSerArgThrAsnAspThrPheAsnIleLysAspIle----- 731  
 QY 1692 TGGATTTCGTTTGTTCGCAACGGACCGTAAAGAGAACACCTATCACATATTCCTGT 1751  
 Db 732 -----SerLeuAsnAspLeuGluLysGluLys----- 741  
 QY 1752 GTGCGAGCCAAACGAAACGAAAGCATGTCTGAATTTCTGAATCTGAAGATGGGA 1811  
 Db 742 -ArgLysLysLysSerGlnHisPheIleAspAsnLeuValLysAlaAspLysAsnG1 760  
 QY 1812 AGTAGAACAG-----CAAACAATATAGTAGTG 1841  
 Db 760 uileSerGluAsnIleAsnLysIleCysAspAsnAsnIleTyrAspGluSe 780  
 QY 1842 CCACATCTGCTGTATTTCCATAGTACCTGCTTACCTCTCCGTCACCAAGAAATGGA 1901  
 Db 780 rIleAsnAsnIleTyr----- 785  
 QY 1902 AGTAGATAGTAAGATGAAGAGGATCCGATGGCTAAGAGAAACCATACAAAT 1961  
 Db 786 -----AspGluSerIleAsnAsnI1 792  
 QY 1962 TGAAGAGTTTCTGTGTTTAAATGAAGAGAGAGAAAGTGTGAACCTCTCGAATCTCCA 2021  
 Db 792 eTyrAspGluSerIleAsnAsnIleTyrAspGluSerIleAsnAsnIleTyrAspGluAs 812  
 QY 2022 TGTCTAGACGATGGTTTATGTGCAATCAAAATGAATGATCTGCTGTATGCTGTGTGT 2081  
 Db 812 nileAsnAsn-----IleTyrAspGluAsnIleAsnAsn-----IleTyrAs 826  
 QY 2082 AGAAAT-----TATGGACAGAAATATTAAGAGAAATTTATGTCGAACTT 2129  
 Db 826 pGluAsnIleAsnAsnIleTyrAspGluGlyIleAsnLys-----IleCys----- 841  
 QY 2130 CATGCTTCATCTAGTCAGCATGCTTAAATCTTATTAGCATTAATGTCAATAGATAA 2189  
 Db 842 -----AspAspAsnIleLeuGluAsnLysAsnIleLysTh 853  
 QY 2190 AGCT-----GTTACCAAGCTCGTGAATCGACGCAAAATTAGAAAGGGGAAATCTGC 2243  
 Db 853 rThrAsnAspIleValGlnValGluAsnAsnGluSerIleGluLysAsnGluLeuMe 873  
 QY 2244 TTCCTCTGCAACGAAGAAATAACTGAAGAACAAATGGACAGCAATGATGATTTAGTGA 2303  
 Db 873 tIleSerLeuAsnLysAspIle-----AsnAsnThrTyrAsnMetPheLysG1 889  
 QY 2304 A-----ATTAACCTAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGG 2348  
 Db 889 uAsnValAspIlePheIleAsnLys-----IleLysArgGluSerLeuLeuLys 905  
 QY 2349 GTTTCAAAACAGACCAAAACAAACTCTGAAAGCTCTAACCCCATCTTATGGACA 2408  
 Db 905 sIleAspLys-AsnIleLysAspAsnAsnAspAsp----- 917  
 QY 2409 AACACTGAAATTACATTTTAGGGAATTCATCTCTAAGAATATGTTTGTGTTTAAATC 2468  
 Db 918 -----GluTyrIleMetAspAsnPheTyrGluAsnAsp-----PheIleIleAsnH 933

QY 2469 ATATGTTCCAA 2479  
 Db 933 iLysMetGlu 936  
 RESULT 13  
 A45555  
 glutamate rich protein - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jun-2000  
 C:Accession: A45555; S27831  
 R:Berre, M.B.; Driegiel, M.; Hogh, B.; Petersen, E.; Rieneck, K.; Riley, E.; Meis, J.F.;  
 J. Mol. Biochem. Parasitol. 49, 119-131, 1991  
 A:Title: Primary structure and localization of a conserved immunogenic Plasmodium falcip  
 rate life cycle.  
 A:Reference number: A45555; MUID:92131041; PMID:1775153  
 A:Accession: A45555  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1271 <BOR>  
 A:Cross-references: EMBL:MS9706; NID:g160311; PID:g160312  
 A:Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBIIP:77802)  
 Alignment Scores:  
 Pred. No.: 0.00321 Length: 1271  
 Score: 154.00 Matches: 150  
 Percent Similarity: 32.77% Conservative: 123  
 Best Local Similarity: 18.01% Mismatches: 270  
 Query Match: 2.00% Indels: 290  
 DB: 2 Gaps: 40

US-09-874-162A-7 (1-4409) x A45555 (1-1271)  
 QY 142 GAGCAGATCGAGGACACACATCGATACAGATCCACGGGTTTGAAG----- 189  
 Db 258 GluHisLeuLysAspGluLysIleGluAsnGluProLeuValHisGluAsnLeuSerIle 277  
 QY 190 -----AAACAAGATTTACAGCAGCAGCAACCTATCTGCCCTGAGTTACATA 234  
 Db 278 ProAsnAspProIleGluGlnIleLeuAsnGlnPro----- 289  
 QY 235 AATAGATTTCATGACATGCTGCCCGCAGAGAGGAGTCCCTAAAGAGAGATTTCAG 294  
 Db 290 -----GluGlnGluThr-----AsnIleGln 296  
 QY 295 CCGAAGCTTCGCTGACTGTGTCAGCTCAGTGTCTCGAGGGAATGTGTCACCTCCCCA 354  
 Db 297 GluGlnLeuTyrAsnGluLysGlnAsnValGluLysGlnAsnSerGlnIlePro--- 315  
 QY 355 CGCCACAGCAGTGAAGCCCTTACTCCCGCGTCAGCCACCCATCACCCTCTCTTCA 414  
 Db 316 -----SerLeuAspLeuLysGluProThrAsnGluAspIleLeuProAsn----- 330  
 QY 415 TTCGCGAGCAGCTCCGACAGAGCAACACAGATCTATAGATTTCTTGAATCGGAAT 474  
 Db 330 ----- 330  
 QY 475 CTCTAGCACCATAATTTTGGCAGAGACTCTTACTTACATGTCTCTCGAAACTCCAGA 534  
 Db 331 -----HisAsnProLeuGluAsnIleLysGlnSerGluSerGlu 343  
 QY 535 ACAAAC---ATCAAAAGGAAACATTTAAAGTTGATGATATGTTTCAAAAGTAGAGAA 591  
 Db 344 IleAsnHisValGlnAspHisAlaLeuProLysGluAsnIleIleAspLysLeuAspAsn 363  
 QY 592 ATGAAGGA-----GAGCAGAGATCTCATAGTGTCTGAGTCTCATTTGAGCTTAGCTTT 645  
 Db 364 GlnLysGluHisIleAspGlnSerGlnHisAsnIleAsnVal---LeuGlnGluAsnAsn 382  
 QY 646 ACTGTTTCTTCCACAAAATGATGACCATCAACAACTCAAGAAAT---GAACAAAT 702  
 Db 383 IleAsnAsnHisGlnLeuGluProGlnGluLysProAsnIleGluSerPheGluProLys 402



```
QY 703 TCTGTACCTCGAAGTCTCTGTTGAAAGTTTGCCACAAAAAAGAAAGATGTAAGT 762
Db      |||:|||||
AsnIleAspSerGluIleIleu----- 410
QY 763 TGTCCAATAAGCAAGTTCACACAGGTAAAGCAGGTCCTTTGATTCTCGTACCTCAAT 822
Db      |||:|||||
ProGluAsnValGluThrGluGluIleIleAspAspValPro 424
QY 823 CAACAAACCGGAAATTTCCCGTCCCTTCAGTTTCCAGTTTCCAGTAATGAAATTTGAA- 876
Db      |||:|||||
SerProLysHisSerAsnHisGluThrPheGluGluGluThrSerGluSerGluHisGlu 444
QY 877 -----CCTAGTAACAGCCAT---ATGGTGAAGTCTTACTCGTTGCTCTATTAGA 921
Db      |||:|||||
GluAlaValSerGluLysAsnAlaHisGluThrValGluHisGluGluThrValSerGln 464
QY 922 GTGACTCTCCAGGAAGAGAGCTTTAATGAATGATTATATGAGAAACCAATGAAAT 981
Db      |||:|||||
GluSerAsnProGluLysAlaAspAsnAspGlyAsnValSerGlnAsnSerAsnGlu 484
QY 982 ATTGATGTCAATGAAGAGCTTCACGACCAAGAAACGAAATCGTGAGGATGGGAAAG 1041
Db      |||:|||||
LeuAsnGluAsnGluPheValGluSer--GluLysSerGluHisGluAlaGlu-- 502
QY 1042 ACATTTGTTGCACAAATCACAGTATTTCATAAACACAGCCGCTTACAGCTTTAGATGGG 1101
Db      |||:|||||
AsnGluGluSerSerLeuGluGluGly 511
QY 1102 GAATATGAA-----GTAGCATGTCAGGAATGAAGAATGTCCAATAAGCAAGAAAGA 1155
Db      |||:|||||
HisHisGluGluIleValProGluGlnAsnGluGluSerGlyGluSerLys----- 529
QY 1156 GCAACATGGGACATTTCTTGAT-----GGGAAGAGCGTCCCTCCATTCGAA 1203
Db      |||:|||||
LeuValAspAsnAspGluGlyPheGluGluAlaHisGlu 544
QY 1204 ACATTTTCTCAGGACCTACGTTGCGAGTTCCTCTCGTTGGCAGGAGACCAAT--- 1260
Db      |||:|||||
AsnPheSerSerGluValSer-----AsnSerGluLeuAsnGlu 557
QY 1261 -----GATAAATCT---ACGGCTCCTATT----- 1281
Db      |||:|||||
AsnGluPheValGluSerAspLysSerValThrGluProAlaGluHisGluGluValVal 577
QY 1282 -----GCCAACTCTGCCACTAGAAATTCAGAGATCTCCATCAGGAAACAAAG 1332
Db      |||:|||||
SerGluGluSerAsnProGluProAlaGluAsnGluGluSerSerIleGluGluAlaHis 597
QY 1333 CTGTGTTCAAGTTAACTTACTCAAACTATTGCTGTAAAGAAATCAATTCAGTACATCTA 1392
Db      |||:|||||
GlnGluGluIleValProGluGlnAsn-----AspGluGluSerGlyGluSerGlyLeu 615
QY 1393 CAACACAGAAAGAAAGAT-----ACTCGAATGAA 1425
Db      |||:|||||
ValAspAsnGluGluGlyAspPheGluGluProAsnHisGluGluPheGluProAspGln 635
QY 1426 AACCCGACAAATTAAGAATATTTTATCAGTTTCTTATAACAACAATACAGAGCAACAA 1485
Db      |||:|||||
AsnAspSerGluLeu-----SerGluAsnGluLeuValGlu 647
QY 1486 ACTGAAGCAAGAGATGACTCGATTCGCTGGTGTACTGTGAATCGCCGCAAACTTTAT 1545
Db      |||:|||||
SerGlu-----LysSerValSer 653
QY 1546 AGTTTACTCAACATCTTAACTCGCATAGCAGATTTATCTTCAACTATGTTTATCAT 1605
Db      |||:|||||
GluProAlaGluHisValGluIleValSerGluLysSerValSerGluProAlaGluHis 673
QY 1606 CCAAAAGGTGTAGATAGATGTTTCTATCAATGAGTGTATTATGATGCTCTATGACGGA 1665
Db      |||:|||||
ValGluIle-----ValSerGluLysSerThrSerGluProAlaGlu 687
```

```
QY 1666 AATCTCTCAGGATATTATCGCCAACTCGATTTGCTTTTAGT-----CGCAACGGACCA 1719
Db      |||:|||||
HisValGluSerValSerGluGlnSerAsnAsnGluProSerGluLysLysAspGlyPro 707
QY 1720 GTTAAGAGAACCTATCACACATATT-----CTTGTGTGCGAGCCCAACGACAAACAA 1773
Db      |||:|||||
ValProSerLysProPheGluGluIleGluLysValAspValGlnProLysIleValAsp 727
QY 1774 GCAAGCATCTCTGAA-----TTTCTTGAATCTGAAGATGGGGAAGTAGAACGACAAAGA 1827
Db      |||:|||||
LeuGlnIleIleGluProAsnPheValAspSerGlnProAsnProGlnGlu----- 744
QY 1828 ACATATAGTAGTGGCCACCAATCGTCTGTATTTCATAGTATGATACCTGCTCTCTCGT 1887
Db      |||:|||||
-----ProValGlu 747
QY 1888 CCACAAGAAATGGAAGTAGAT-----AGTGAAGATGAAAAG-----GAT 1926
Db      |||:|||||
ProSerPheValLysIleGluLysValProSerGluGluAsnLysHisAlaSerValAsp 767
QY 1927 CTTGAATGCTTAAGAAAAAACCAATTACACAATTTGAAGAGTTCCTTCGAT----- 1977
Db      |||:|||||
ProGluValLysGluLysGluAsnValSerGluValValGluGluLysGlnAsnSerGln 787
QY 1978 -----GTTAATGAAGGAGAGAAAGAGTGTATGATAAATCTGGAAT 2016
Db      |||:|||||
GluSerValGluGluIleProValAsnGluAspGluPheGluAspValHisThrGluGln 807
QY 2017 CTCCATGTCTAGCAGCATGGTTTATTGCTGCAATCAATCAATCAATCAATCAATCAAT 2076
Db      |||:|||||
Leu-----AspLeuAspHis----- 812
QY 2077 TTGTAGAAAAATTTGGACAGAAATTAATTAAGAGAAATTTATGCGAAACTTCATGCTT 2136
Db      |||:|||||
----- 812
QY 2137 CATCTAGTCAGCATCGATGACTTTAATCTTATTAGCATAATGTCAATAGATAAAGCTGTT 2196
Db      |||:|||||
-----LysThrValAspProGluIle 819
QY 2197 ACAAGCTCCGTGAAATGCAGCAAAAATAGAAAAGGGGAAATCTGCTTCCCTGCAAC 2256
Db      |||:|||||
ValGluValGluGluIleProSerGluLeuHisGluAsnGluValAlaHisProGluIle 839
QY 2257 GAAGAAATAACTGAA-----GAACAAAATGGGACAGCAAAATGGAATTTAGT 2301
Db      |||:|||||
ValGluIleGluGluValPheProGluProAsnGlnAsn-----AsnGluPheGln 856
QY 2302 GAAATTAATCTAAAAGAGAAAGCT-----TTGGAACACAGATAGT 2340
Db      |||:|||||
GluIleAsnGluAspLysSerAlaHisIleGlnHisGluIleValGluValGluGlu 876
QY 2341 GTCTCAGGGGTTTCAAAACAGACAGCAAAAACAAAACCTC 2379
Db      |||:|||||
LeuProGluAspAspLysAsnGluLysVal 887
```

## RESULT 14

A46302

PTB-associated splicing factor, long form - human

N;Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding pr

C;Species: Homo sapiens (man)

C;Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text\_change 24-Sep-1999

C;Accession: A46302; A43557; S29995

R;Fattori, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.

Genes Dev. 7, 393-406, 1993

A;Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.

A;Reference number: A46302; MUID:93194059; PMID:8449401

A;Accession: A46302

A;Molecule type: mRNA

A;Residues: 1-707 4P4T&gt;

A;Cross-references: EMBL:X70944; NID:G38457; PID:CAA50283.1; PID:G38458

A;Note: sequence extracted from NCBI backbone (NCBI:127206)

R;Gower, H.J.; Moore, S.E.; Dickson, G.; Elson, V.L.; Nayak, R.; Walsh, F.S.



Development 105, 723-731, 1989  
A:Title: Cloning and characterization of a myoblast cell surface antigen defined by 24.1  
A:Reference number: A43557; MUID:90091812; PMID:2480677  
A:Accession: A43557  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 312-707 <GOW>  
A:Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712  
C:Genetics:  
A:Gene: GDB:SFPO; PSF  
A:Cross-references: GDB:138275  
A:Map position: 4q-4q  
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
C:Keywords: alternative splicing; pre-mRNA splicing; surface antigen  
F:298-359/Domain: ribonucleoprotein repeat homology <RRM1>  
F:372-438/Domain: ribonucleoprotein repeat homology <RRM2>

Alignment Scores:  
Pred. No.: 0.0035 Length: 707  
Score: 153.00 Matches: 120  
Percent Similarity: 29.95% Conservative: 60  
Best local Similarity: 19.97% Mismatches: 207  
Query Match: 1.99% Indels: 214  
DB: 2 Gaps: 22

US-09-874-152A-7 (1-4409) x A46302 (1-707)

QY	2	CCGCCCGCGCTCGCAGACCGCAGCACCACCGGGGCTCTCGATGTAGCACCATCACAGGCA	61
Db	35		
QY	62	TCGCGCGCGCTCTCTTCTCCAACTCCGATTCGGGGCTCGGGACTCCACTTCC	121
Db	55		
QY	122	CCACCTTGGCGACTCATCGACACATCGAGGACACCATCATGATCCACGGG	181
Db	74		
QY	182	TTTTAGAAAACAAAGATTACAGACCACTTGTGCCCTGAGTTACATAATAGAT	241
Db	87		
QY	242	TCATGACAGATGTCGCGCGCAGACGAGTCCCTTAAGAGAAGATTACGCCGACG	301
Db	93		
QY	302	TCTCGCTGACTCTGTCCAGCTCACTGTCGAGGAATGTGCCACTCCCGCAGGCACA	361
Db	109	oVal-ValAlaGlnGlyProGlyProAlaProGlyValGlySerAlaProProAlaSerS	129
QY	362	GCATGGAAGCTTACTTCCCGGTGACCCACCCATCACC	415
Db	129	erSer-----AlaProProAlaThrProThrSerGlyAlaProProGlySerG	146
QY	416	TCCGACGACGACTCCGACAGAGCCACACAGATCTATAGATTCTTCGAACTCGGAATC	475
Db	146	lyProGlyProThrProThrProProAlaVal	157
QY	476	TCATAGACCAATATTTTGCACAGAACTTACTTACATGTCTCATCGAAACTCCAGAA	535
Db	157	-----	157
QY	536	CAACATCAAAAGGAAACATTTAAAGTTGATGATGTTATCAAAAGTAGAGAAATGA	595
Db	157	-----	157
QY	596	AAGGAGACGAAGAATCTCATAGCTTGTTCAGCTTACGTTTACTGTTTCT	655
Db	158	-----	161
QY	656	TCCACAAAATGATGAAGCCATCCAAATCAGAAATGACAAATCTGTACCTGG	715
Db			

RESULT 15  
 S69972  
 Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N  
 N:Alternate names: protein N2453  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 09-Mar-1996 #sequence\_revision 06-Sep-1996 #text\_change 20-Jun-2000  
 C:Accession: S69972  
 R:Berger, P.; Doignon, F.; Crouzet, M.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S62975  
 A:Accession: S69972  
 A:Molecule type: DNA  
 A:Residues: 1-1749 <BER>  
 A:Cross-references: EMBL:Z71330; NID:G1301918; PID:CAA95924.1; PID:G1301920  
 A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4  
 C:Genetics:  
 A:Map position: 14L  
 A:Mobile element: retrotransposon Ty1.N  
 C:Superfamily: TyB protein

Alignment Scores:  
 Pred. No.: 0.0047 Length: 1749  
 Score: 152.00 Matches: 177  
 Percent Similarity: 35.59% Conservative: 118  
 Best Local Similarity: 21.35% Mismatches: 264  
 Query Match: 1.98% Indels: 270  
 DB: 2 Gaps: 48

US-09-874-162A-7 (1-4409) x S69972 (1-1749)

QY	163	ATCGATACAGATCCACGGGTTTAA-----GAAACAAAGAAATTACAGCAGCCCAACC	213
DB	881	ValAspThrThrAsnTyrValIleLeuGlnGlyLysGluSerArgLeuAspGlnPheAsn	900
QY	214	TATGTGCTCCAGTAC-----ATAAATAGATTCATGACAGAT-----	252
DB	901	TyrAspAlaLeuThrPheAspGluAspLeuAsnArgLeuThrAlaSerTyrGlnSerPhe	920
QY	253	---GCTGCCCCGAGACGAGGAGTCCCTAAGAGAGAGATTCAGCCGAGCTCTCCGCTG	309
DB	921	IleAlaSerAsnGluIleGlnSerAspAspLeuAsnPheGlnSerAspIleGluLeu	940
QY	310	-----ACTGTCCAGCTCAGTGTCTCAGGAGAAATGCTCCACT	348
DB	941	HisProGluGlnProArgAsnValLeuSerLysAlaValSerProThrAsp---SerThr	959
QY	349	CCCCCA-----CGCCACGACGTGGAGCCCTTACTCCCCCC	384
DB	960	ProProSerThrHisThrGluAspSerLysArgValSerLysThrAsnIleArgAlaPro	979
QY	385	-----GTGACCCCAACCCATCCACCCCTCC-----TCTTCATCCGACGACG	426
DB	980	ArgGluValAspProAsnIleSerLysSerAspIleLeuProSerLysLysArgSerSer	999
QY	427	ACTCCGACAGACCAACACAGATCATAGATTTCTCGAATCCGAAT-----	474
DB	1000	ThrPro-----GlnIleSerAspIleGluSerThrGlySerGlyGlyMethHis	1015
QY	475	-----CTCATGACCAATATTTTGGACAGACTCTTACTTACATGCTCT	519
DB	1016	ArgLeuAspValProIleLeuAlaProMetSerGlnSerAsnThr-----	1030
QY	520	CATCGAAACTCCAGAACAAACATCAAAAGGAAACATTAAGATTGATGATGATCA	579
DB	1031	HisGluSerSerHisThrSer---LysSerLysAspPheArgHisSerAspSerTyrSer	1049
QY	580	-----GlnIleSerAspIleGluSerThrGlySerGlyGlyMethHis	1015
DB	1050	AspAsnGluThrAsnHisThrAsnValProIleSerSerThrGlyGlyThrAsnAsnLys	1069
QY	583	GTAGAGAAATGAAAGGAGACGAAGATCTCATAGCTTGTGAGCTCATTTGCAGCTTACG	642

DB	1070	ThrValProGlnThrSerGluGlnGluThrGluLysArgIleIleHis-----	1085
QY	643	TTTACTGGTTCTTCCACAAAATGATAGCCATCACCACAACTCAGAAAATGAACAAAAT	702
DB	1086	-----ArgSerProSerIleAspThrSerSerSerGluSer	1097
QY	703	TCTGTACCTCGAAGTCTCTGTGAAA-----GTTTCCACAAAAAAGAAAG	753
DB	1098	AsnSerLeuHisHisValProIleLysThrSerAspThrCysProLysGluAsnThr	1117
QY	754	GATGTAGTTGTCCATTAAGCAAGTTCCTCCACAGTAAAGACAGGTGCTTTGATTCCT	813
DB	1118	GluGluSerIle---IleAlaAspLeuPro-----LeuPro	1128
QY	814	GACCTCAATCAACAAACAAACCCGAAATTTCCGCTCCCTTCGAGTTCCAGTAATGAATTT	873
DB	1129	AspLeuProGluProPro-----ThrLysLeuSerAspSerPheLysGluLeu	1145
QY	874	GAACCTAGTAAACAGCCATATGTTGAAGTCTTACTGCTGCTTATTTAGAGTACTGCTCA	933
DB	1146	ProIleAsnSerArgGlnThrAsnSer-----	1155
QY	934	GGAAGAGAGAGTTTAATGGAATGATTAAATGAGAAACCAATGAAATATTGATGTCAAT	993
DB	1156	-----SerLeuGlyGlyIle-----GlyAspSerAsnAlaTyrThrIleAsn	1170
QY	994	GAAAGAGCTTCCAGCCAGAGAAACGAAATCOTGAGGATGGGAAAGACATTT---GTT	1050
DB	1171	Ser-----LysLysArgSerLeuGluAspAsnGluThrGluIleLysVal	1185
QY	1051	GCACAAATGACAGTATTTGATTAACAAACAGCGCTTACAGCTTTAGATGGGAAATATGA	1110
DB	1186	SerArgAspThrTrpAsnThrLysAsnMetArg-----	1196
QY	1111	GTAGCCATCGAGAAATGGAAGATGTCCAATTAAGCAAGAAAGAAAGAA---GCA	1158
DB	1197	-----SerLeuGluProProArgSerLysLysArgIleHisLeuIleAla	1211
QY	1159	ACATGGAGACTATTTCTGATGGAGAGAGGCTCCATTCGAAACATTTTCTCAGGGA	1218
DB	1212	AlaValLysAlaVal-----LysSerIleLysProIleArgThr-----	1224
QY	1219	CCTAGTTCAGTTC-----ACTCTGCTGTCGACGAGACCAATGATAAAATCTACG	1272
DB	1225	-----ThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAspIleLysGluLysGlu	1242
QY	1273	GCTCCTATTGCCAAACCTCTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAG	1332
DB	1243	-----LysTyrIleGluAlaTyrHisLysGlu-----	1251
QY	1333	CCTGTTTCAAGTTAAACCTACTCAAACTATTGCTGTTAAAGATCATTTGACTACAGATCTA	1392
DB	1252	-----ValAsnGlnLeuLeuLysMetLysThrTrpAspThrAspLysTyr	1266
QY	1393	CAAAACAAGAAAGAAAGGATCTCCAAATGAAACCCGACAAAAATTAAGAAATTTTAT	1452
DB	1267	TyrAspArgLysGluIleAsp-----ProLysArgValIleAsnSer	1280
QY	1453	CAGTTTCTCTATAACCAATACA-----AGCAACAAACCAACCAACAGAT	1500
DB	1281	MetPheIlePheAsnArgLysArgAspGlyThrHisLysAlaArgPheValAlaArgGly	1300
QY	1501	GACCTCAATTCGCT-----TGGTGTACTCTGCACTCCGCGCAAACTTTAT	1545
DB	1301	AspIleGlnHisProAspThrTyrAspSerGlyMetGlnSerAsnThrValHisTyr	1320
QY	1546	AGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTATCTTCACTATCTTTATCAT	1605
DB	1321	AlaLeuMetThrSerLeuSerLeuAlaLeuAspAsn-----AsnTyrTyrIle---	1336
QY	1606	CCAAAGGTGCTAGGATAGATGTTTCTATCAATGATGTTTATGCTGCTTCCTCAGCA	1665
DB	1337	-----ThrGlnLeuAspIleSer-----SerAlaTyr-----LeuTyrAlaAsp	1349





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 25, 2004, 18:16:23 ; Search time 76 Seconds

(without alignments)

6041:508 Million cell updates/sec

Title:

US-09-874-162A-7

Perfect score:

7690

Sequence: 1 cccgcggcgctgcagag.....tacataaagtccaataatat 4409

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09874162/runat\_25082004\_171843\_2169/app\_query.fasta\_1.4551  
-DB=swissprot42 -Qfmt=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874162 @C@N 1 1 18 @runat 25082004\_171843\_2169 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3413	44.4	739	1 JUAZ_HUMAN	Q15022 homo sapien
2	3332	43.3	741	1 JUAZ_MOUSE	Q80u70 mus musculus
3	1127.5	14.7	900	1 SU12_DROME	Q9njg9 drosophila
4	661	8.6	243	1 JAZ1_HUMAN	Q86vz6 homo sapien
5	661	8.6	243	1 JAZ1_MOUSE	Q80zq5 mus musculus
6	224	2.9	631	1 ENP2_ARATH	Q816y4 arabidopsis
7	217.5	2.8	440	1 VN22_ARATH	Q8w5b1 arabidopsis
8	211	2.7	692	1 FIS2_ARATH	Q9znt9 arabidopsis
9	176.5	2.3	1142	1 GIN4_YEAST	Q12263 bos taurus
10	164.5	2.1	1453	1 Y373_BOVIN	Q8tu23 bos taurus
11	161	2.1	1875	1 MLP1_YEAST	Q02455 saccharomyc
12	159.5	2.1	912	1 UP3_YEAST	Q01477 saccharomyc
13	157	2.0	1208	1 PCP1_SCHPO	Q92351 schizosacch
14	155.5	2.0	1755	1 YU27_YEAST	Q27098 saccharomyc
15	153	2.0	707	1 SPQ_HUMAN	P23246 homo sapien
16	152	2.0	1755	1 YU29_YEAST	P47100 saccharomyc
17	152	2.0	1957	1 SPOF_SCHPO	Q10411 schizosacch
18	151	2.0	1328	1 YNT5_YEAST	Q04214 saccharomyc

19	149.5	1.9	2867	1 RBP2_PLAVB	Q00799 plasmodium
20	149	1.9	2869	1 RBP1_PLAVB	Q00798 plasmodium
21	148.5	1.9	2230	1 GOA4_HUMAN	Q13439 homo sapien
22	147.5	1.9	1790	1 USO1_YEAST	P25386 saccharomyc
23	147	1.9	1085	1 YAP2_SCHPO	Q09863 schizosacch
24	147	1.9	1328	1 YMD9_YEAST	Q03434 saccharomyc
25	146.5	1.9	1235	1 TRK1_YEAST	P12685 saccharomyc
26	145	1.9	1744	1 TANA_XENLA	Q01550 xenopus lae
27	145	1.9	1953	1 BN11_YEAST	P41832 saccharomyc
28	145	1.9	5560	1 SPEN_DROME	Q8ex83 drosophila
29	143.5	1.9	698	1 YB06_YEAST	P38283 saccharomyc
30	143	1.9	2349	1 TPR_HUMAN	P12270 homo sapien
31	141	1.8	1275	1 RGA2_SCHPO	Q10164 schizosacch
32	140.5	1.8	5430	1 MACF_HUMAN	Q9upn3 homo sapien
33	140.5	1.8	5938	1 MAC4_HUMAN	Q96pk2 homo sapien
34	139.5	1.8	1833	1 ZEP2_HUMAN	P31629 homo sapien
35	139	1.8	1328	1 YME4_YEAST	Q04711 saccharomyc
36	139	1.8	3210	1 CENF_HUMAN	P49454 homo sapien
37	138	1.8	1328	1 YMU0_YEAST	Q04670 saccharomyc
38	138	1.8	1972	1 BA2B_HUMAN	Q9uif8 homo sapien
39	137	1.8	774	1 RIN1_RAT	P97680 rattus norv
40	137	1.8	1969	1 Z292_HUMAN	Q60281 homo sapien
41	136	1.8	1544	1 ARHC_HUMAN	Q9nzn5 homo sapien
42	135.5	1.8	836	1 UME6_YEAST	P39001 saccharomyc
43	134.5	1.7	1148	1 RFC1_HUMAN	P35251 homo sapien
44	133.5	1.7	2663	1 CENE_HUMAN	Q02224 homo sapien
45	133	1.7	2245	1 MYSU_DICDI	P54697 dictyosteli

#### ALIGNMENTS

RESULT 1

JUAZ\_HUMAN

ID JUAZ\_HUMAN STANDARD; PRT; 739 AA.

AC Q15022; Q96BD9;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Polycarb protein SUZ12 (Suppressor of zeste 12 protein homolog)

DE (joined to JAZF1 protein) (Chromatin precipitated E2F target 9

DE protein) (CHET 9 protein).

GN SUZ12 OR JUAZ1 OR CHET9 OR KIAA0160.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI\_TaxID=9606;

RN [1]

RX SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RX MEDLINE=96127530; PubMed=8590280;

RT "Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1.";

RL DNA Res. 2:167-174(1995).

RN [2]

RX SEQUENCE FROM N.A.

RC TISSUE=Uterus;

RX MEDLINE=22388257; PubMed=12477932;

RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Valek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Murry D.M., Sodergren B.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shvchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,



Db 312 AspGlyGluTyrGluValAlaMetGlnGluMetGluGluCysProIleSerIleValSerGly 331  
 QY 1156 GCAACATGGAGACTATCTTCTGATGGAGAGCGCTCCCTCATTCGAAACATTTCTCAG 1215  
 Db 332 AlaThrTrpGluThrIleLeuAspGlyArgLeuProPheGluThrPheSerGln 351  
 QY 1216 GGACCTACCTTTCAGTTCACTCTCTGTTGGACGAGGACCAATGATAATCTACGGCT 1275  
 Db 352 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspIleSerThrAla 371  
 QY 1276 CTTATTCGCAACCTCTGCTCCACTAGAAATTCAGAGAGTCTCCATCAGAGAAACAGCCT 1335  
 Db 372 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 391  
 QY 1336 GGTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCATTCACACAGATCTCAA 1395  
 Db 392 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrAspLeuGln 411  
 QY 1396 ACAAGAAAAGAAAGATCTCCAAATGAAACCCGACAAAATTAAGAAATTTATATCAG 1455  
 Db 412 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGln 431  
 QY 1456 TTTCTCTATAACAACATACAGCAACCAACTGAGCAAGAGATGACTGCATTCGCCCT 1515  
 Db 432 PheLeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspLeuHisCysPro 451  
 QY 1516 TGGTGTACTCTGAATCGCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT 1575  
 Db 452 TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysCysHis 471  
 QY 1576 ACCAGATTATCTTCAACTATGTTTATCATCAAAAGTGTAGATAGATGTTCTTATC 1635  
 Db 472 SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle 491  
 QY 1636 AATCAGTGTATGATGGCTCCATCGAGGAATCTCAGGATATTATCATCGCAACTCGGA 1695  
 Db 492 AsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly 511  
 QY 1696 TTTGTTTATGTCGCAACGAGCATGTTAAGAACACCTCATCAACATATCTTGTGTGC 1755  
 Db 512 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 531  
 QY 1756 AGGCCAAACGACAAAGCAGCATCTGATTTCTTGATCTGAATCGAAGATGGGGAAGTA 1815  
 Db 532 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 551  
 QY 1816 GAACAGCAAAACACATAGTAGTGGCCACCAATCGTCTGTTTCCATAGTATGATCCTGC 1875  
 Db 552 GluGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerAspThrCys 571  
 QY 1876 TTACTCTCTCCGTCACAGAAATGGAAGTAGATAGTAGAAGTAAAGATCTCGAATGG 1935  
 Db 572 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 591  
 QY 1936 CTAAGAGAAAACCATACACAAATTCGAAGAGTTTCTGATGTTTAAAGAGGAGAGAAA 1995  
 Db 592 LeuArgGluLysThrIleThrGlnIleGluPheSerAspValAsnGluGlyGluLys 611  
 QY 1996 GAAGTGATGAACACTCGGAATCTCATGATGATGATGATGATGATGATGATGATGATGAT 2055  
 Db 612 GluValMetLysLeuThrAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 631  
 QY 2056 ATGATCATCGCTGTATCTGTTCTGAGAAATATGACAGCAAAATTAATTAAGAGAT 2115  
 Db 632 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysAsn 651  
 QY 2116 TTATCTGAAACTTCATGCTTTCATCTAGTCAGCATGCTGATCTTAACTTATAGCATA 2175  
 Db 652 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 671  
 QY 2176 ATGTCATAGATAAGCTGTACCAAGCTCGTGAATTCGACGAAATTAATTAAGAAAGGG 2235

Db 672 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 691  
 QY 2236 GAATCTCTCCCTCCCTGCAAAACGAGAAATTAACAGAGAAACAAATGCGACAGCAATGGA 2295  
 Db 692 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 711  
 QY 2296 TTTAGTCAAAATTAACCTCAAAAGAGAAAGCTTTGAAACACAGATAGTGTCTCAGGGTTTCA 2355  
 Db 712 PheSerGluLeuAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 731  
 QY 2356 AACAGAGCAAAAACAAAAAATCTC 2379  
 Db 732 LysGlnSerLysLysGlnLysLeu 739  
 RESULT 2  
 JJAZ\_MOUSE  
 ID JJAZ\_MOUSE STANDARD; PRT; 741 AA.  
 AC Q80U70; Q80Y10; Q99L07;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Polycomb protein Suz12 (Suppressor of zeste 12 protein homolog).  
 GN SUZ12 OR KIAA0150 OR D11ERTD530E.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RC MEDLINE=22579291; PubMed=12693553;  
 RX Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,  
 Nakajima D., Nagase T., Ohara O., Koga H.;  
 RA "Prediction of the coding sequences of mouse homologues of KIAA gene:  
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous  
 RT cDNAs identified by screening of terminal sequences of cDNA clones  
 RT randomly sampled from size-fractionated libraries.";  
 RL DNA Res. 10:35-48(2003).  
 RN [2]  
 SEQUENCE OF 365-741 FROM N.A.  
 RP STRAIN=FVB/N; TISSUE=Kidney;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butlerfield F.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by  
 CC forming multiprotein complexes, which are required to maintain the  
 CC transcriptionally repressive state of homeotic genes throughout  
 CC development. PcG proteins are not required to initiate repression,  
 CC but to maintain it during later stages of development. They  
 CC probably act via the methylation of histones, rendering chromatin  
 CC heritably changed in its expressibility. Component of the PRC2  
 CC complex, which methylates lys-9 and lys-27 residues of histone H3  
 CC (by similarity).  
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,  
 CC SUZ12/JJAZ1, RBBP4 and RBBP7. This complex is distinct from the

CC PRC1 complex, which contains many other PcG proteins like Bmi1,  
CC CBX4, CBX8, PHC3, PHC9, SCMH1, RING1 and RNF2 (By similarity).  
CC -- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -- SIMILARITY: Belongs to the VEPs (VRN2-EMF2-FIS2-SU(2)12) family.  
CC -- SIMILARITY: Belongs to the VEPs (VRN2-EMF2-FIS2-SU(2)12) family.  
CC -- SIMILARITY: Contains 1 C2H2-type zinc finger.  
CC -- CAUTION: Ref.1 sequence differs from that shown due to an  
CC erroneous RNA joining and initiations.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL curation -  
CC European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC	EMBL; AK122213; BAC65495.1; ALT_SEQ.
DR	EMBL; BC003922; AAH03922.1; -
DR	EMBL; BC051099; AAH51099.1; -
DR	MGI; 12611758; D11ExtD30e.
DR	InterPro; IPR007087; znf C2H2.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; FALSE NEG.
KW	Transcription regulation; Repressor; Chromatin regulator;
KW	Nuclear protein; Metal-binding; Zinc; Zinc-finger.
FT	ZN FING 450 473 C2H2-TYPE.
FT	DOMAIN 565 641 VFES-BOX.
FT	DOMAIN 7 50 GLY-RICH.
FT	DOMAIN 29 69 ALA-RICH.
SO	SEQUENCE 741 AA; 83025 MW; 4A6B6E52D24FAEF2 CRC64;

Alignment Scores:		
Pred. No.:	6.3e-227	Length: 741
Score:	332.00	Matches: 630
Percent Similarity:	98.61%	Conservative: 9
Best Local Similarity:	97.22%	Mismatches: 9
Query Match:	43.33%	Indels: 0
DB:	1	Gaps: 0
US-09-874-162A-7 (1-4409) x JJA2 MOUSE (1-741)		

916	Qy	TTTAGAGTGA	CTCGTC	CAGGACAA	GAGAGTGTTAAATGG	AATGATTAATCGA	AGAACCAAT	975
254	Dd	PheArgValThr	ArgProGly	ArgArgGluPhe	AsnGlyMetIle	AsnGlyGluThr	Asn	273
976	Qy	GAATAATTGAT	GTCAAT	GAAGACTTCC	CAGCACAAGA	AAAAACGA	AATCCTGAGGATGGG	1035
274	Dd	GluAsnIleasp	ValSerGlu	GluLeuPheAla	ArgArgGlySer	IleArgAsnArgGlu	AspGly	293
1036	Qy	GAAGAAGACAT	TTCCTGC	CAAAATGAC	AGTATTGTGATA	AAAAACAGGCGCTT	TACAGCTTTTA	1095
294	Dd	GluLysThrPhe	ValAlaGln	MetThrValPhe	AspLysAsnArg	ArgLeuGlnLeuLeu		313
1096	Qy	GATGGGGAATAT	CAAGTAGC	CATGAGCA	GAATGGAA	GAATGCCAAT	TAACAAGAAAAGA	1155
314	Dd	AspGlyGluTy	rGluValAla	MetGlnLeuMet	GluGluCysPro	IleSerLysLysArg		333
1156	Qy	GCAACATGGG	AGACTATTCT	TGTGGA	GAGAGCTCCCT	CCAATCGAA	CAATTTTCTCAG	1215
334	Dd	AlaThrTrp	GluThrIleLeu	AspGlyLysArg	LeuProPhe	GluThrPheSer	Gln	353
1216	Qy	GGACCTACGT	TTCAGTTC	CTCTTCGT	TGCACAGG	AGACC	CAATGATAAATCTACGGCT	1275
354	Dd	GlyProThr	LeucGlnPhe	ThrLeuArgTrp	ThrGlyLeuThr	AsnAspLysSer	ThrAla	373
1276	Qy	CCATTATGCC	AAAACTCT	TGCCCAC	TAGAAATTC	CAGAGAGTCT	CCATCAGAAAAACAAGCCT	1335
374	Dd	ProValAla	LysProLeu	AlaThrArg	SerSerGluSer	LeuHisGlnGlu	AsnLysPro	393
1336	Qy	GGTTCAGTTAA	CCCTACT	CAAACTT	TGCTGTAA	AGAATCAT	TGACTACAGATCTACAA	1395
394	Dd	GlySerVal	LysProAla	GlnThrIleAla	ValLysGluThr	LeuThrThrGlu	LeuGln	413
1396	Qy	ACAAGAAAAGA	AAAGATACT	CCCAAATG	AAAAACCG	AAAAATTA	AGAATATTTTATCAG	1455
414	Dd	ThrArgLys	GluLysAsp	AsnSerAsnGluSer	ArgGlnLysLeu	ArgIlePhe	TyrGln	433
1456	Qy	TTTCTCTATA	CAACAATACA	AGCGCA	CAAACTGA	AGCAAGAGAT	CCTCGAATGCCCT	1515
434	Dd	PheLeuTy	rAsnAsnThrArg	GlnGlnThrGlu	AlaArgAsp	AspLeuHisCysPro		453
1516	Qy	TGCTGTACT	CTGACG	CGCAACTT	TAGTTTACT	CAAGCATCTT	AACTCTGCCAT	1575
454	Dd	TrpCysThr	LeuAsnCys	ArgLysLeuTy	SerLeuLeuLys	HisIleuLysLeu	CysHis	473
1576	Qy	AGCAGATTTAT	CTTCAACT	ATGTTTAT	CTCAAAAG	GTGCTAGGATAGAT	TTTTCTATC	1635
474	Dd	SerArgPhe	IlePheAsnTy	rValTy	HisProLysGly	AlaArgIleAsp	ValSerIle	493
1636	Qy	AATGAGTGT	TATGATGG	CTCTATG	CAGGAAATC	CTCAGGATAT	TTCATCGCCAACTGGA	1695
494	Dd	AsnGluCys	tyrAspGlySer	Ty	rAlaGlyAsnPro	GlnAspIleHis	ArgGlnProGly	513
1696	Qy	TTTGCTTTT	AGTCGCA	ACGACAGT	TAAGAGAA	CACTTACACATAT	TCTGTGTGC	1755
514	Dd	PheAlaPhe	SerArgAsn	GlyProVal	LysArgThrPro	IleThrHisIleLeu	ValCys	533
1756	Qy	AGGCCAAAAC	GAACAAAG	CAATGTCT	GGAATTTCT	GAAATCTCA	AGATGGGGAAGTA	1815
534	Dd	ArgProLys	ArgThrLysAla	SerMetSerGlu	PheLeuGluSer	GluAspGlyGlu	Val	553
1816	Qy	GRACACGA	AGACATAT	AGTAGT	GTGGCCCAAT	CTGCTGTATTT	TCCATAGTGA	1875
554	Dd	GluGlnGln	ArgThrTy	SerSerGlyHis	AsnArgLeuTy	rPheHisSer	AspThrCys	573
1876	Qy	TTTACTCT	CTCCCA	GAAGAAAT	TGGAAGTAG	ATGTCAGAT	GAAGGATCTCTGAATGG	1935
574	Dd	LeuProLeu	ArgProGln	MetGluVal	AspSerGlu	AspGluLysAsp	ProGluTrp	593
1936	Qy	CTAAGAGAAA	AAACCAT	TACACAAAT	TGAAGT	TTTTCTGATG	TTTAAATGAGGAGAAA	1995
594	Dd	LeuArgGlu	LysThrIleThrGln	IleGluGluPhe	SerAspVal	AsnGluGlyGlu	Lys	613
1996	Qy	GAAGTGAT	GAATCAAT	CTCTGGAAT	CTCCCATG	TCATGA	AGCATGGGTTTATGTCTGACAATCAA	2055



```

Db      614  GluValMetCysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 633
QY      2056  ATGAATCATGCTGTATGCTGTTGTAGAAAATTATGGACAGAAATAATTAAGAAGAAAT 2115
Db      634  MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysLysAsn 653
QY      2116  TTATGTCGAACTTCATGCTTCATCTAGTCAGCATGCATCTTAACTTATTAGCATATA 2175
Db      654  LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 673
QY      2176  ATGTCAATAGATAAAGCTGTATACCAAGCTCCGTGAAATGCAGCAAAAATATGAAAAGGGG 2235
Db      674  MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGlnLysGly 693
QY      2236  GAATCTGCTTCCCTCCGAAACGAGAAATACTAGGAACAAATAATGGACAGCAAAATGGA 2295
Db      694  GluSerAlaThrProSerAsnGluGluIleAlaGluGluGlnAsnGlyThrAlaAsnGly 713
QY      2296  TTTAGTGAATTAACCTCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGGTTTCA 2355
Db      714  PheSerGluThrAsnSerLysLysAlaLeuGluThrAspGlyValSerGlyValPro 733
QY      2356  AAACAGACAGCAAAAACAAAATC 2379
Db      734  LysGlnSerLysLysGlnLysLeu 741
RESULT 3
SUI2:DRONE
ID      SUI12:DRONE      STANDARD;      PRT;      900 AA.
AC      Q9NUG9; Q8T9D8; Q9VW55;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Polcomb protein Su(z)12 (Suppressor 12 of zeste protein).
GN      Su(z)12 OR CG8013.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT
RX      MEDLINE=21430867; PubMed=11546753;
RA      Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,
RA      Rasmuson-Lestander A., Mueller J.;
RT      "Su(z)12, a novel Drosophila Polcomb group gene that is conserved in
RT      vertebrates and plants.";
RL      Development 128:3371-3379 (2001).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=Berkley;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA      Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA      Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

```

RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Mizny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E.R., Stadling A.C., Stapleton M., Strong R., Sun E.,
RA      Swirskas R., Testor C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195 (2000).
RN      [3]
RP      REVISIONS, AND ALTERNATIVE SPLICING.
RX      MEDLINE=2445069; PubMed=12537572;
RA      Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA      Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA      Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA      Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA      Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA      Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA      Lewis S.E.;
RT      "Annotation of the Drosophila melanogaster euchromatic genome: a
RT      systematic review.";
RL      Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
RN      [4]
RP      SEQUENCE FROM N.A. (ISOFORM 2).
RX      STRAIN=Berkley; TISSUE=Embryo;
RA      Scapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA      George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA      Rubin G.M., Celniker S.E.;
RT      "A Drosophila full-length cDNA resource.";
RL      Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
RN      [5]
RP      IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAF1 AND ESC,
RP      AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX      MEDLINE=22296673; PubMed=12408863;
RA      Czermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V.;
RT      "Drosophila enhancer of Zeste/ESC complexes have a histone H3
RT      methyltransferase activity that marks chromosomal Polycomb sites.";
RL      Cell 111:185-196 (2002).
RN      [6]
RP      IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAF1 AND ESC, AND
RP      METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX      MEDLINE=22296674; PubMed=12408864;
RA      Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,
RA      Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;
RT      "Histone methyltransferase activity of a Drosophila Polycomb group
RT      repressor complex.";
RL      Cell 111:197-208 (2002).
CC      -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
CC      forming multiprotein complexes, which are required to maintain the
CC      transcriptionally repressive state of homeotic genes throughout
CC      development. PcG proteins are not required to initiate repression,
CC      but to maintain it during later stages of development. They
CC      probably act via the methylation of histones, rendering chromatin
CC      heritably changed in its expressibility. Component of the Esc/E(z)
CC      complex, which methylates Lys-9 and Lys-27 residues of histone H3.
CC      Despite the presence of a zinc-finger, it does not bind directly
CC      to DNA, the Esc/E(z) complex being probably recruited to DNA by
CC      Pto. The Esc/E(z) complex is necessary but not sufficient to
CC      recruit a functional PcG repressive complex that represses target
CC      genes, suggesting that the recruitment of the distinct PRC1
CC      complex is also required to allow a subsequent repression.
CC      -!- SUBUNIT: Component of the Esc/E(z) complex, composed of Esc, E(z),

```

CC Su(z)12, Rpd3, Caf1 and probably Pho. This complex is distinct  
 CC from the PRC1 complex, which contains many other PcG proteins like  
 CC Pc, Ph, Psc, Su(z)2. The two complexes however cooperate and  
 CC interact together during the first 3 hours of development to  
 CC establish PcG silencing.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=B;  
 CC IsoId=Q9NUG9-1; Sequence=Displayed;  
 CC Name=2; Synonyms=A;  
 CC IsoId=Q9NUG9-2; Sequence=VSP\_007033, VSP\_007034;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Belongs to the VFPs (VRN2-BMF2-FIS2-SU(Z)12) family.  
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL; AF149047; AAF73149.1; -;  
 CC DR EMBL; AE003515; AAF49094.2; -;  
 CC DR EMBL; AE003515; AAN11641.1; -;  
 CC DR EMBL; AY069809; AAL39954.1; -;  
 CC DR FLYBase; FBgn0020887; Su(z)12.  
 CC DR GO; GO:0000790; C:nuclear chromatin; NAS.  
 CC DR GO; GO:0003677; F:DNA binding; NAS.  
 CC DR GO; GO:0016458; P:gene silencing; NAS.  
 CC DR GO; GO:0016481; P:negative regulation of transcription; NAS.  
 CC DR InterPro; IPR007087; Znf\_C2H2.  
 CC DR SMART; SM00355; Znf\_C2H2; 1.  
 CC DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 CC DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
 CC KW Transcription regulation; Repressor; Developmental protein;  
 CC Nuclear protein; Metal-binding; Zinc; Zinc-finger;  
 CC Alternative splicing.  
 CC FT ZN\_FING 411 434 C2H2-TYPE.  
 CC FT DOMAIN 527 603 VEF5-BOX.  
 CC FT DOMAIN 355 366 ASN-RICH.  
 CC FT DOMAIN 699 879 SER-RICH.  
 CC FT VARSPPLIC 806 855 NTVLNKKQRYSDGSPGTGIGNHGCGSGSGANRNKSNHSL  
 CC FT PATSNASS -> VEQADAPVETHSDNADVGIIDDECG  
 CC FT GFGAVGVNGVSPVANNVCVN (in isoform 2).  
 CC FT FTId=VSP\_007033.  
 CC FT Missing (in isoform 2).  
 CC FT FTId=VSP\_007034.  
 CC FT G->D: IN SU(Z)12-2; INDUCES LARVAL  
 CC FT LETHALITY WHEN HOMOZYGOUS.  
 CC SQ SEQUENCE 900 AA; 100104 MW; 53BA0D83C49EC92F CRC64;  
 CC -----  
 CC Alignment Scores:  
 CC Pred. No.: 1.06e-71 Length: 900  
 CC Score: 1127.50 Matches: 273  
 CC Percent Similarity: 49.46% Conservative: 137  
 CC Best Local Similarity: 32.93% Mismatches: 240  
 CC Query Match: 14.66% Indels: 179  
 CC DB: 1 Gaps: 26  
 CC -----  
 CC US-09-874-162A-7 (1-4409) x SU12\_DROME (1-900)  
 CC QY 244 ATGACAGATGCTGCCCGGACGACGAGTCCCTAAGAGAGAGATTCAGCCGAGCTC 303  
 CC Db 1 MetAlaProAlaLysLysArgGluLysAspSer----- 11  
 CC QY 304 TCGCTGACTCTGTCCAGCTCAGTGTCTCTCAGGAGGAATGTG-----TCCACTCCC 351  
 CC Db 12 -----AsnProAspGlySerAlaAlaAsnGlyLlelleGlyLeuThrHisGlyAlaPro 29  
 CC QY 352 CCAGCCACACAGCTGGNAGGCTTACTTCCCTCCCTGACCCCAACCCATCCACCCCTCTCT 411

Db 30 AspAlaSerAsnAlaGlySerThrValProThrAlaGluGlyGlnValLysLeuAsn 49  
 QY 412 TCATCCGCGACGACACTCCG-----ACAGAGCCACACAGATCTAT 453  
 Db 50 GlyHisGlnGlnGlnGlnGlnPheLeuGlnAlaPheGluLysProThrGlnLleTyr 69  
 QY 454 AGATTTCCTTCGAACCTCGGAATCTCATAGCACCAATATTTTTCACAGAACTCTTACTTAC 513  
 Db 70 ArgTyrLeuArgAsnArgHisGluThrAsnProLlePheLeuAsnArgThrLysSerTyr 89  
 QY 514 ATGTCTCATCGAAATCTCCAGAACACATCAAAAGGAAACATTTAAAGTTGATCATATG 573  
 Db 90 MetLysGluArgMetSerArgAsnAsnLysLysArgLleSerPheGlnValAsnSerMet 109  
 QY 574 TTATCAAAAGTAGAGAAATGAAGAGAGCAAGAAATCTCATAGCTTGTGAGCT---CAT 630  
 Db 110 LeuGluSerIle-----ThrGlnLysSerGluAlaValSerGlnAsnTyr 124  
 QY 631 TTCAGCTTACGTTTACTGTTTCTTCCACAAATAATGATAGCCATCACCAACTCAGAA 690  
 Db 125 LeuHisValIleTyrAspSerLeuHisGluLysLeuProAlaArgLeuAspAsnGluSer 144  
 QY 691 AATGACAA-----AATTCGTGTACCTGAA 717  
 Db 145 GlyGluAspLeuLeuGlnGlnLeuLeuCysGluAlaGlyGluSerValSerValGlu 164  
 QY 718 GTCCTGCTTGTGAAAGTTTGCACAAAAAGAGATGTAAGTTGTCCATTAAGGCAA 777  
 Db 165 ThrThrLeuTyrLysIleThrArgSerLysArgLysAspSerThrLeuAspPheGlnGlu 184  
 QY 778 GTTCCACAGGTAAAGACAGGTGCTTGTGATCTCTGACCTCAATCAACAAACCCCGA 837  
 Db 185 LeuLeuSerLysCysSerGlnIleValTyrAsnPro-----LysAspArgValGly 201  
 QY 838 AATTTCCCGTCTGTCAGTTTCCAGTAATGAATTTGAACCT---AGTAACAGCATATG 894  
 Db 202 GluHisAlaThrIleSerIleProLeuGlnThrMetArgProMetGlyGluGlnHisThr 221  
 QY 895 GTGAAGCTTACTCTGCTTACTATTAGATGACTCTGTCAGGAGGAAGAGAGATTAAATGA 954  
 Db 222 Leu-----TyrLysLeuLeuPheArgIleLys----- 230  
 QY 955 ATGATTATGAGAGAACCAATGAAATATTGATGTCATGAAGAGCTTCCAGCCAGAGA 1014  
 Db 231 ValLeuSerProSerThrCysAsn-----AspGluAsnAlaGluThrProPro---Asn 247  
 QY 1015 AAACGAAATCGTGAAGATGGGAAAGACATTTGTGCACAAATGACAGTATTGTATATAA 1074  
 Db 248 LysArgSerArgProAsn---GluLysMetPheGlySerGluLeuLleLeuTyrGluLys 266  
 QY 1075 ACACGGCTTACAGCTTTAGATCGGGAATATGAAGTAGCCATGAGCAAGAAATGGAAGA 1134  
 Db 267 SerSerGlyPhe---IleThrGluGlyGluTyrGluAlaMetLeuGlnProLeuAsnSer 285  
 QY 1135 TGTCCAATA-----AGCAAGAAAGAGCAACATGGGAGACTATTCTTGATGGGAAG 1185  
 Db 286 ThrSerIleLysSerPheSerProLysLysCysThrTyrGluThrMetProAsp---Ser 304  
 QY 1186 AGCTGTCCT-----CCATTGAAACATTTCTCAGGACCTACCTTCAGTTCAGTTCCT 1236  
 Db 305 TyrIleProLeuSerLeuThrTyrAspValTyrGlnGlnSerProMetLeuLysPheHis 324  
 QY 1237 CTTCTGTGACAGGAGAGACCAATGATAAATCTACGGCTCTCTATTCGCAACCTCTTCC 1296  
 Db 325 LeuThrLeu-----SerAsnGlnGlnLeuProGluMetIleSerAlaProGluLeu 341  
 QY 1297 ACTAGA-----AATTCAGAGAGTCTC 1317  
 Db 342 GlnArgTyrValGlnHisLeuAspAlaValAlaGluMetAsnTyrAsnAsnAsnTyr 361  
 QY 1318 CATCAGGAAAC-----AAGCTGTCTCAGTTAAACCTACTCAACTATT 1362

Db 362 AsnAsnAsnAsnCysSerGlyLeuLysAsnGlySerGlyGlyAsnSerThrVal 381  
 QY 1363 GCTGTTAAAGATCATCTACAGATCTACAAACAAAGAAAGAGATCTCCAAT 1422  
 Db 382 Cys-----LysThrThrPro--- 386  
 QY 1423 GAAACCGACAAAATTAAGAAATATTTATCAGTTCTCTATAACAAACAAATCAAGGAA 1482  
 Db 387 -----GluHisIleGlnIleValTyrAsnPhemetyrSerAsnAsnThrArgGln 403  
 QY 1483 CAACTGAAGCAAGAGATGACCTGCATCCCTTGGTGACTCTGAACTCGCGCAACTT 1542  
 Db 404 GlnThrGluTyrThrGlnGluLeuAsnCysProTyrCysGlyLeuAspCysLeuArgLeu 423  
 QY 1543 TATAGTTTACTCAAGCATCTTAACTCTCCATACAGATTTATCTTCAACTATGTTAT 1602  
 Db 424 TyrAlaLeuLeuLysHisLeuLysLeuCysHisAlaArgPheAsnPhethrTyrGlnPro 443  
 QY 1603 CATCCAAAAGGTCTAGGATAGATGTTCTCATATGAGTGTATGATGGTCTCTATGCA 1662  
 Db 444 AlaGlySerGlyAlaArgIleAspValThrIleAsnAspAlaTyrAspGlySerTyrAla 463  
 QY 1663 GGAATCTCCAGATATTCATCGCAACCTGATTTGCTTTAGTCGCAAC---GGACCA 1719  
 Db 464 GlySerProTyrAspLeuAlaGlyProSerGlySerSerPheAlaArgThrCysGlyPro 483  
 QY 1720 GTTAAGACAAACCTTATCACATATCTTGTCAGCGCCAAACGCAACAAAGCAAGC 1779  
 Db 484 ValArgArgThrSerValThrSerLeuMetValCysArgProArgArgGlnLysThrCys 503  
 QY 1780 ATGCTCTGAATTTCTTGAATCTGAAGATGGGAGTAGAAGCAAGCAAGATATAGTAGT 1839  
 Db 504 LeuAspGluPheLeuGluLeuAspGluAspGluIleSerAsnGlnArgSerTyrIleThr 523  
 QY 1840 GCCCAATCGTCTGTATTCATAGTATGATCTTACCTTACCTCCGTCACAGAAATG 1899  
 Db 524 GlyHisAsnArgLeuTyrHisHisThrGluThrCysLeuProValHisProLysGluLeu 543  
 QY 1900 GAAGTAGATAGTGAAGATCAAGATCTCTGAATGGCTTAAGAGAAAACCATACACAA 1959  
 Db 544 AspileAspSerGluGlyGluSerAspProLeuTyrLeuArgGlnLysThrIleGlnMet 563  
 QY 1960 ATTGAAGATTTCTGATGTTAATGAAGAGAGAAAGATGATGAATCTGGAATCTC 2019  
 Db 564 IleAspGluPheSerAspValAsnGluGlyGlyLysGluLeuMetLysLeuTrpAsnLeu 583  
 QY 2020 CATGTCGAAGCATGGGTTTATGCTGACAAATCAATGAATCATGCTGATCTCTTT 2079  
 Db 584 HisValMetArgHisGlyPheValGlyAspCysGlnLeuProIleAlaCysGluMetPhe 603  
 QY 2080 GTAGAAATATGACAGACAAATTAATGAAGAAATTTATGCGAAATCTCATGCTTCAT 2139  
 Db 604 LeuAspAlaLysGlyThrGluIleValArgLysAsnLeuTyrArgAsnPheIleLeuHis 623  
 QY 2140 CTAGTCAGCATGCTGACTTAATCTTATAGCATAATGTCATAGATAAGCTGTACC 2199  
 Db 624 MetCysSerLeuPheAspTyrGlyLeuIleAlaAlaGluThrValTyrLysThrValGln 643  
 QY 2200 AAGTCCTCGTAATCGACAAAATTAAGAAAGGGGAATCT----- 2241  
 Db 644 LysLeuGlnGlyLeuLeuSerLysTyrAlaAlaGlyGlnGluMetGlnArgGlnArg 663  
 QY 2241 ----- 2241  
 Db 664 GluGluGlnLeuLysTyrTrpLeuAspValGlyMethHisLysLysGlnGluAspProLys 683  
 QY 2242 -----GCTTCCCTCCAAACGAAGAAATAACTGAA----- 2271  
 Db 684 ThrLeuLysSerProGlnLysProAlaProProAlaAspGlnAlaSerThrSerSerAla 703  
 QY 2271 ----- 2271  
 Db 704 SerThrSerGlySerGlySerGlySerSerMetGlnProProLysArgMetProAla 723

QY 2272 -----GAACAAATGGACAGCAAT 2292  
 Db 724 HisLeuLysArgGlySerAlaAlaSerSerProGlyValGlnSerLysGlyThrGluAsn 743  
 QY 2293 GCATTTAGTGAATTAATTAATCTCMAAGAGAAAGCTTTGGAAACAGATAGTCTCAGGGTT 2352  
 Db 744 GlyThrAsnGlySerAsnSer-----SerSerSerAsnSerLysAsnVal 758  
 QY 2353 TCAAAACAGACAGAAAAACCAAAACTC 2379  
 Db 759 AlalysLysSerAlaAspGlnProLeu 767  
 RESULT 4  
 JAZ1 HUMAN  
 ID JAZ1\_HUMAN  
 AC Q86VZ6; Q8N3L7; STANDARD; PRT; 243 AA.  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Juxtaposed with another zinc finger protein 1.  
 GN JAZF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain, and Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallory S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Vallaloon D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,  
 Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Amalgam;  
 RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH SUZ12.  
 RX MEDLINE=21265470; PubMed=11371647;  
 RA Koontz J.I., Soreng A.L., Nucci M., Kuo F.C., Pauwels P.,  
 van Den Berghe H., Cln P.D., Fletcher J.A., Sklar J.;  
 RT "Frequent fusion of the JAZF1 and JAZ1 genes in endometrial stromal  
 tumors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353(2001).  
 CC -!- FUNCTION: Potential transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear protein (Potential).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q86VZ6-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q86VZ6-2; Sequence=VSP\_007755;  
 CC Note=Sequence incomplete. No experimental confirmation  
 available;

CC -!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal  
CC rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21)  
CC chromosomal translocation involving JAZF1 and SUZ12 generates the  
CC JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1  
CC and the C-terminus part of SUZ12. The translocation is frequently  
CC found in all cases of endometrial stromal tumors, except in  
CC endometrial stromal sarcomas, where it is rarer.  
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; BC042441.1; -;  
CC EMBL; AL834234; CAD38912.1; -;  
CC MIN; 606246; -;  
CC InterPro: IPR007087; Znf\_C2H2.  
CC Pfam; PF00096; Znf\_C2H2; 3.  
CC SMART; SM00355; Znf\_C2H2; 3.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2; FALSE\_NEG.  
CC PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; FALSE\_NEG.  
CC Transcription regulation; Nuclear protein; Metal-binding; Zinc;  
CC Zinc-finger; Chromosomal translocation; Proto-oncogene;  
CC Alternative splicing.  
CC FT ZN\_FING 12 37 C2H2-TYPE 1.  
CC FT ZN\_FING 173 198 C2H2-TYPE 2.  
CC FT ZN\_FING 208 230 C2H2-TYPE 3. (DEGENERATE)  
CC FT SITE 129 130 BREAKPOINT FOR TRANSLLOCATION TO FORM  
CC FT VARSPLIC 1 38 JAZF1-SUZ12 ONCOGENE.  
CC FT MTGLAASFFSNTFCGCGLHFTPLADLIEHEDNHI ->  
CC FT YFVLGSLIYDAKSKKRNVCVIT (in isoform 2).  
CC FT /FTId=VSP 007755.  
CC FT D -> G (IN REF. 1).  
CC FT CONFLICT 133 133 27079 MW; 878A6EE3D82C0588 CRC64;  
CC SEQUENCE 243 AA; 27079 MW; 878A6EE3D82C0588 CRC64;

Alignment Scores:  
Pred. No.: 6,3e-39 Length: 243  
Score: 661.00 Matches: 128  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 8.60% Indels: 0  
DB: 1 Gaps: 0

US-09-874-162a-7 (1-4409) x JAZ1\_HUMAN (1-243)  
QY 52 ATGACAGCATCGCGCGCTCTCTCTCCATACCTCGCATCGGGGGCTCGGGA 111  
DB 1 MetThrGlyLeuAlaAlaSerPheSerAsnThrCysArgPheGlyGlyCysGly 20  
QY 112 CTCACACTTCCCGCCGACCTCATCGAGCAGCATCGAGCAACACCATCATACA 171  
DB 21 LeuHisPheProThrLeuAlaAspLeuGluHisLeuGluAspAsnHisLeuAspThr 40  
QY 172 GATCCAGCGTGTAGAAAACAGAAATACAGAGCAGCAACCTATGTGCTCGATTAC 231  
DB 41 AspProArgValLeuGluGlyGlnGluGlnGlnProThrThrValAlaLeuSerTy 60  
QY 232 ATAAATAGATTTCATCAGAGTGTCTCCCGCCGAGAGCAGGAGTCCCTAAAGAAGATT 291  
DB 61 IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysile 80  
QY 292 CAGCCGAAGCTCTCGTACTCTGTCCAGTCTGAGTGTCTCAGGGAATGTTCACATCCC 351  
DB 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100  
QY 352 CCACGCCACAGAGTGGAGGCTTACTCCCGCTGACCCACCATCACCCCTCTCTCT 411  
DB 101 ProArgHisSerSerGlySerLeuThrProProValThrProProLeuThrProSerSer 120

QY 412 TCATTCCGACGACGACATCTCGACA 435  
DB 121 SerPheArgSerSerThrProThr 128  
RESULT 5  
JAZ1\_MOUSE  
ID JAZ1\_MOUSE STANDARD; PRT; 243 AA.  
AC Q802Q5; Q8BLQ8; Q8BWN4;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Juxtaposed with another zinc finger protein 1.  
GN JAZF1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A. (ISOFORM 1)  
RP STRAIN=CS7BL/6J; TISSUE=Brain cortex, and Pancreas;  
RX MEDLINE=22354683; PubMed=12466851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,  
RA Baladrelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konggava A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
RA Ravasi T., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wyszewski-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan A., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 2).  
RP TISSUE=Testicle;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max A.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Raheij J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

-!- FUNCTION: Potential transcription factor.

-!- SUBCELLULAR LOCATION: Nuclear protein (Potential).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

isoId=Q80ZQ5-1; Sequences=Displayed;

Name=2;

isoId=Q80ZQ5-2; Sequences=VSP 007756;

Note=No experimental confirmation available;

-!- SIMILARITY: Contains 3 C2H2-type zinc fingers.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; AK043767; BAC31649.1; -.

EMBL; AK050474; BAC34275.1; -.

EMBL; BC048577; AAH48577.1; -.

InterPro; IPR007087; Znf C2H2.

Pfam; PF00096; zf-C2H2; 3.

SMART; SM00355; Znf C2H2; 3.

PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2.

PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.

Transcription regulation; Nuclear protein; Metal-binding; Zinc; Zinc-finger; Alternative splicing.

FT ZN\_FING 12 37 C2H2-TYPE 1.

FT ZN\_FING 173 198 C2H2-TYPE 2.

FT ZN\_FING 208 230 C2H2-TYPE 3 (DEGENERATE).

FT VARFUNG 64 64 Missing (in isoform 2).

FT FTID=VSP\_007756.

FT CONFLICT 211 211 R -> L (IN REF. 1; BAC34275).

SQ SEQUENCE 243 AA; 27097 MW; 878A6EB3D82C0348 CRC64;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
6.3e-39	661.00	243	128
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.60%	Indels:	0
DB:	1	Gaps:	0

US-09-874-162A-7 (1-4409) x JAZ1\_MOUSE (1-243)

QY 52 ATGACAGCATCCCGCCGCTCTTCTTCTCCAAATACCTGCGGATTCGGGCTGCGGA 111

Db 1 MetThrGlyIleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly 20

QY 112 CTCGACTTCCACCTGCGCCACCTCATCGAGCATCGAGGCAACACCATCATGATACA 171

Db 21 LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr 40

QY 172 GATCCACGGGTTTATAGAAAAAACAAGATTACAGCAGCAACCTATGTTGCCCTGAGTTAC 231

Db 41 AspProArgValLeuGlyGlnGlnGlnGlnProThrTyValAlaLeuSerTyr 60

QY 232 ATAAATAGATTATGACAGATCTCCCGCCGAGAGCGAGAGTCCCTAAAGAGAGATT 291

Db 61 IleAsnArgPheMetThrAspAlaAlaArgArgGluGlnGluSerLeuLysLysLysIle 80

QY 292 CAGCCGAGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 351

Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100

QY 352 CCACGCCACAGAGTGGAGCTTACTCCCGGTCGACCCACCCACCCACCCCTCTCTCT 411

Db 101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSerSer 120

QY 412 TCATTCCGACGAGCAGCTCCGACA 435

Db 121 SerPheArgSerSerThrProThr 128

RESULT 6

EMF2\_ARATH STANDARD; PRT; 631 AA.

AC Q8L6Y4; Q93V59; Q9LU50; Q9LU51;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Polycomb protein EMBRYONIC FLOWER 2.

GN EMF2 OR AF5GS1230/AF5GS1240 OR MWD22.18/MWD22.19.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_taxID=3702;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC STRAIN=cv. Columbia;

RX MEDLINE=21558794; PubMed=11701882;

RA Yoshida N., Yanai Y., Chen L., Kato Y., Hiratsuka J., Miwa T.,

RT "Sung Z.R., Takahashi S.;

RT "EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates

RT shoot development and flowering in Arabidopsis.";

RL Plant Cell 13:2471-2481 (2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20181125; PubMed=10718197;

RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,

RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence

RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC

RT clones.";

RL DNA Res. 7:31-63 (2000).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC STRAIN=cv. Columbia;

RX MEDLINE=22954850; PubMed=14593172;

RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Cheuk R.P.,

RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.P.,

RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,

RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,

RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,

RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,

RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,

RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,

RA Khan S., Kosemura E., Ishida J., Jiang P.X., Jones T., Kawai J.,

RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,

RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,

RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;

RT "Empirical analysis of transcriptional activity in the Arabidopsis

RT genome.";

RL Science 302:842-846 (2003).

-!- FUNCTION: Polycomb group (PcG) protein. Involved in flowering

processes by repressing unknown target genes and preventing

reproductive development. PcG proteins act by forming multiprotein

complexes, which are required to maintain the transcriptionally

repressive state of homeotic genes throughout development. PcG

proteins are not required to initiate repression, but to maintain

it during later stages of development. They probably act via the

methylation of histones, rendering chromatin heritably changed in

its expressibility.

-!- SUBUNIT: In plants, PcG complexes are probably composed of a

member of the EZ family (CLF or MEA), FIE, and a member of the

VEFS family (FIS2, VRN2 or EMF2) (By similarity).

-!- SUBCELLULAR LOCATION: Nuclear (Probable).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

isoId=Q8L6Y4-1; Sequence=Displayed;

Name=2;

```

CC CC Isold-Q816Y4-2; Sequence=VSP_007456;
CC CC Note=No experimental confirmation available;
CC CC -1- TISSUE SPECIFICITY: Widely expressed throughout the life cycle
CC CC with higher levels in proliferating tissues. Expressed in both
CC CC vegetative and the reproductive shoot meristems.
CC CC -1- DEVELOPMENTAL STAGE: Expressed in the developing embryos and
CC CC endosperm, then decreases when embryos mature and soon after
CC CC cellularization in the endosperm. After germination, it is
CC CC expressed in the shoot apical meristems (SAMs), leaf primordia,
CC CC and young leaves. In the reproductive shoots, it is expressed in
CC CC both the inflorescence and floral meristems. Later, it is
CC CC expressed in floral organ primordia. In cofilin, it is
CC CC expressed in SAMs and lateral organs. In roots, it is expressed in
CC CC root tips.
CC CC -1- SIMILARITY: Belongs to the VEPs (VRN2-EMP2-FTS2-SU(Z)12) family.
CC CC -1- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC CC gene model prediction.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AB053171; BAB58956.1; -
CC CC EMBL; AB053262; BAB58957.1; -
CC CC EMBL; AB023044; BAA97386.1; ALT-SEQ.
CC CC EMBL; AB023044; BAA97387.1; ALT-SEQ.
CC CC EMBL; AY140086; AAW98227.1; -
CC CC TRANSFAC; T05378; -
CC CC InterPro; IPRO07087; Znf_C2H2.
CC CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
CC CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE NEG.
CC CC Transcription regulation; Repressor; Flowering; Nuclear protein;
CC CC Zinc; Zinc-finger; Metal-binding; Alternative splicing.
CC CC ZN FING 324 347 C2H2-TYPE.
CC CC FT DOMAIN 505 583 VEFs-BOX.
CC CC FT DOMAIN 32 35 POLY-GLU.
CC CC FT DOMAIN 289 292 POLY-SER.
CC CC FT DOMAIN 396 399 POLY-ARG.
CC CC FT VARSPLIC 392 396 Missing (in isoform 2).
CC CC FT SEQUENCE 631 AA; 71680 MW; F36FE92D0F62E610 CRC64;
CC CC -----
CC CC Alignment Scores:
CC CC Pred. No.: 4.17e-08 Length: 631
CC CC Score: 224.00 Matches: 132
CC CC Percent Similarity: 33.82% Conservative: 102
CC CC Best Local Similarity: 19.08% Mismatches: 252
CC CC Query Match: 2.91% Gaps: 206
CC CC DB: 1
CC CC -----
CC CC US-09-874-162A-7 (1-4409) x EMF2_ARATH (1-631)
CC CC QY 436 GAGCCAAACAGATCATATGTTATCTTCGAACTCGGAATCTCATAGCACCAATATTTTG 495
CC CC Db 47 LysProValGluLeuTyAsnIleIleGlnArgArgAlaIleArgAsnProLeuPheLeu 66
CC CC QY 496 CACAGAACTCTTACTTCATGCTTCATTCGAACTCCAGAACCAACATCAAAAGGAACA 555
CC CC Db 67 GlnArgCysLeuHisTyLysIleGluAlaLysHisLysArgArgIleGlnMetThrVal 86
CC CC QY 556 TTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAAATGAAGAGAGCAAGAATCTCAT 615
CC CC Db 87 PheLeu-----SerGly 90
CC CC QY 616 AGCTTGTGAGCTCATTTGCGAGCTTACGTTTACTGCTTTCTTCACAAAAATGATAAGCCA 675
CC CC Db 91 AlaIleAsaAlaGlyValGln----- 97

```

676 TCACCAAACTCAGAAAATGACAATAATTCTGTGTACCCTCGGAAGTCTCTGCTTGTAAGTT 735  
Db -----ThrGlnLysLeuPheProLeuTyrlleuleuleualaArgLeu 111  
Qy TGCCACAAAAAAGGAAGATGTAAGTTGTCCTCAATAAGGCAA----- 777  
Db ValSerProLys-----ProValAlaGluTyrserAlaValTyArg 125  
Qy -----GTTCCACAGCATAAAGACAGGTGCCT----- 804  
Db PheSerArgAlaCysIleLeuThrGlyGlyLeuGlyValAspGlyValserGlnAlaGln 145  
Qy -----TTGATTCTCGACTCAAT-----CAACAAAAACCCGGAAAT--- 840  
Db AlaasnPhelLeuLeuProaspMetasnArgLeuAlaLeuGluAlalysSerGlySerLeu 165  
Qy -----TTCCCGTCCTTCGAGTTTCCAGTTAATGAATGAACCTAGTAGAACGCCAT 891  
Db AlaIleLeuPheIleSerPheAlaGlyAlaGlnAsnSerGluPheGlyIleaspSerGly 185  
Qy ATGCTGAAGCTTACTCTGTTGCTATTATTAGAGTCACCTCGTCCAGGAAGAAGAGTTTTAAT 951  
Db LysIleHisSer-----SerLeuTyralaserTrpGln 211  
Qy GGAATGATTAAATGAGAAACCAATGAAAAATTATGATGTCAATGAAGAGTTCGAGCCAGA 1011  
Db GlyAsnIleGlyGlyHiScys-----LeuTrpSerLysIleProLeuGln 204  
Qy AAAAAAGAAATCGTGAGATGGGAAAAAGACATTTGTTGCACAAATGACAGATTGTAT 1071  
Db -----SerLeuTyralaserTrpGln 211  
Qy AAAAAACAGCGCTTACAGCTTTTATAGTGGGAATATGAAGTAGCCATGCAGGAAATGGAA 1131  
Db LysSerProasnMetaspLeuGlyGlnArgValasphThrValserLeuValGluMetGln 231  
Qy GAATGTCCAATAGCAAGAAAAAGACAGCATGGGAGACTATTCTTGCATGGGAAGAGCGTG 1191  
Db ProCysPheIleLysLeuLysSerMetSerGluGluLysCysValSer----- 247  
Qy CCTCCATTGCAACATTTCTCAGGACCTACTCTTCAGTTCACCTCTCGTTGGACAGGA 1251  
Db -----IleGlnValProSerAsnProLeuthr----- 256  
Qy GAGACCAATGATAAATCTACGGCTCTATTGCCCCAACCTCTTGCCCACTAGAAATTCAGAG 1311  
Db SerSerProGlnGlnValGlnValThrIleSerAlaGluGlu 271  
Qy AGTCTCCATCAGAAAAACAGCGTGTTCAGTTAAACCTACTCAAACTATTGCTGTATAA 1371  
Db ValGlyserThrGluLysSerProtyrSerPheSerTyraSnAspIleSer---Ser 290  
Qy GAATCATTTACTACAGATCTCAACACAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1431  
Db SerSerLeuLeuGlnIleIleargLeuArg-----ThrGlyAsn----- 303  
Qy CAAAAATTAAGAAATATTTATACGTTTCTCTCTATATACAAATACAGGCAACAAACTGAA 1491  
Db ValValPheAsnTyra-gTyTyraSnAsnLysLeuGlnLysThrGlu 319  
Qy GCAAGAGATGACCTGCATTCGCTCGTGTACTCTGAACCTGCCGCAAACTTTATATGTTTA 1551  
Db valThrGluaspPheSerCysProPheCysLeuValLysCysAlaSerPheLysGlyLeu 339  
Qy CTCACGATCTTAAACTCTGCCATACAGATTTATCTTCAACTATGTTTATCATCCAAAA 1611  
Db ArgTy-HisLeuProserThrhHisaspLeuAsnPheGluPheTrpValThrGluGlu 359  
Qy GTGTGTAGATAGATGTTTCTATCAATGAGTGTATGATGGTCTCTATCGAGAAATCCT 1671  
Db PheGlnAlaValAsnValSerLeuLysThsrGluThrMetIleSerLysVal---AsnGlu 378  
Qy CAGGATATTCAAT---CGCCAACTCGATTTGCTTTT----- 1704

Db 379 AspAspValAspProGlnGlnThrPhePhePheSerSerLysLysPheArgArg 398  
 QY 1705 -----AGTCGCAACGGACCA----- 1719  
 Db 399 ArgGlnLysSerGlnValArgSerArgGlnGlnProHisLysLysLeuGlyCysGlu 418  
 QY 1720 -----GTTAAGAGAACCTATACATATCTTGTCGAGGCCAAAACGA----- 1767  
 Db 419 ValLeuAspLysThrAspAlaHisSerValArgSerGluLysSerArgIleProPro 438  
 QY 1767 ----- 1767  
 Db 439 GlyLysHisTyrGluArgIleGlyAlaGluSerGlyGlnArgValProProGlyThr 458  
 QY 1768 -----ACAAAGCA 1776  
 Db 459 SerProAlaAspValGlnSerCysGlyAspProAspTyrValGlnSerIleAlaGlySer 478  
 QY 1777 ACATGCTGCAATCTTCAATCTGAAGATGGGAGTAGACACAA-----AGACAA 1830  
 Db 479 ThrMetLeuGlnPheAlaLysThrArgLysIleSerIleGluArgSerAspLeuArgAsn 498  
 QY 1831 TATAGTAGTGGCCACAATCGTCTGATATTCATAGTAGTACCTGCTTACCTTCGCTCCA 1890  
 Db 499 ArgSerLeuLeuGlnLysArgGlnPhePheHisSerHisArgAlaGlnProMetAlaLeu 518  
 QY 1891 CAAGAAATG-----GAAGTAGATGAGATGAAGATGATCGATCGCTAAGAGAA 1944  
 Db 519 GluGlnValLeuSerAspArgAspSerGluAspGluValAlaAspValAlaAspPhe 538  
 QY 1945 AAAACCATACACAAATCAAGATTTCTGATGTTAATGAAGGAGAGAAAGATGATG 2004  
 Db 539 GluAspArgArgMetLeuAspAspPheValAspValThrLysAspGluLysGlnMetMet 558  
 QY 2005 AAATCTGGAATCTCATGATGATGAGCATGGTATTGTCGACAAATCAATGAATCAT 2064  
 Db 559 HisMetTrpAsnSerPheValArgLysGlnArgValLeuAlaAspGlyHisIleProTrp 578  
 QY 2065 GCCTGATGCTCTTCTAGAAAATATGACAGAAAATAATTAAG---AAGATTTATGT 2121  
 Db 579 AlaCysGluAlaPheSerArgLeuHisGlyProIleMetValArgThrProHisLeuLe 598  
 QY 2122 CGAAACTTCATGCTTCATCTAGTCAGCATGCTTAAATCTTATAGCATAATGCA 2181  
 Db 599 TrpCysTrpArgValPheMetValLysLeuTrpAsnHisGlyLeuLeuAspAlaArgThr 618  
 QY 2182 ATAGATAAGCTGTACCAAGCTCCGTGAAATGCAG 2217  
 Db 619 MetAsnAsnCysAsnThrPheLeuGlnLeuGln 630  
 RESULT 7  
 ID VRN2\_ARATH STANDARD; PRT; 440 AA.  
 AC Q8W5B1; O23524; Q8W5B2; Q94CF5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Polycarb group protein VERNALIZATION 2.  
 GN VRN2 OR AT4G16845 OR FCAL1.23 OR DL4450W.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OC NCBI\_TaxID=3702;  
 RN [1]\_TaxID=3702;  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), FUNCTION, AND CHARACTERIZATION.  
 RC STRAIN=cv. Landsberg erecta;  
 RX MEDLINE=21575875; PubMed=11719192;  
 RA Gendall A.R., Levy Y.Y., Wilson A., Dean C.;  
 RT "The VERNALIZATION2 (VRN2) gene mediates the epigenetic regulation of  
 RT vernalization in Arabidopsis.";  
 RL Cell 107:525-535 (2001).

[2] SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RC MEDLINE=98121113; PubMed=9461215;  
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,  
 RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,  
 RA Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,  
 RA Wedler H., Wedler E., Wambutt R., Weitzinger T., Pohl T., Terryn N.,  
 RA Gielen J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,  
 RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,  
 RA Kouter P., Entian K.-D., Rieger M., Schaefer M., Funk B.,  
 RA Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,  
 RA Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P.,  
 RA Piravandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,  
 RA Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansoorge W.,  
 RA Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,  
 RA Klosterman S., Schueller C., Chalwatzis N.;  
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of  
 RT Arabidopsis thaliana";  
 RL Nature 391:485-488 (1998).  
 [3] SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RC MEDLINE=20083489; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohelsel A., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J., Vandenbussche F.,  
 RA van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzinger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Moolijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Ganderath K., Dauner D., Herzl A.,  
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenot O., Quigley P., Clabaud G., Muendlein A., Feilner R.,  
 RA Schnabl S., Hillier R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chefidor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Laetelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,  
 RA Nelson J., Spiech J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,  
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana";  
 RL Nature 402:769-777 (1999).  
 [4] SEQUENCE FROM N.A. (ISOFORM 1).  
 RP STRAIN=cv. Columbia;



RX MEDLINE=22954850; PubMed=14593172;  
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.W., Akiyama K., Ansari Y.,  
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
RA Hayashizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M.,  
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Empirical analysis of transcriptional activity in the Arabidopsis  
RL genome.";  
CC -!- FUNCTION: Polycomb group (PcG) protein. Plays a central role in  
CC -!- functionalization by maintaining repressed the homeotic gene FIC, a  
CC floral repressor, after a cold treatment. PcG proteins act by  
CC forming multiprotein complexes, which are required to maintain the  
CC transcriptionally repressive state of homeotic genes throughout  
CC development. PcG proteins are not required to initiate repression,  
CC but to maintain it during later stages of development. They  
CC probably act via the methylation of histones, rendering chromatin  
CC heritably changed in its expressibility.  
CC -!- SUBUNIT: Probable component of a PcG complex. In plants, PcG  
CC complexes are probably composed of a member of the EZ family (CLF  
CC or MEA), FIE, and a member of the VEFs family (FIS2, VRN2 or ERF2)  
CC (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q8W5B1-1; Sequence=Displayed;  
CC Name=2; Synonyms=VRN2;  
CC IsoId=Q8W5B1-2; Sequence=VSP\_007457, VSP\_007458;  
CC -!- TISSUE SPECIFICITY: Weakly expressed. Expressed both during, and  
CC in the absence of vernalization.  
CC -!- SIMILARITY: Belongs to the VEPs (VRN2-EMF2-FIS2-SU(Z)12) family.  
CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to  
CC erroneous gene model prediction.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF284500; AAL32135.1; -;  
CC EMBL; AF284501; AAL32136.1; -;  
CC EMBL; 297342; CAB10457.1; ALT\_SEQ.  
CC EMBL; AL161545; CAB80955.1; ALT\_SEQ.  
CC EMBL; AY034902; AAK59409.1; -;  
CC EMBL; AY063047; AAL34221.1; -;  
CC InterPro; IPR007087; Znf C2H2.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
CC PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; FALSE NEG.  
CC Transcription regulation; Repressor; Nuclear protein; Zinc;  
CC Zinc-finger; Metal-binding; Alternative splicing.  
CC ZNF FING 86 111  
CC C2H2-TYPE.  
CC DOMAIN 156 163  
CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC DOMAIN 267 345  
CC VEFs-BOX.  
CC DOMAIN 398 402  
CC POLY-ASN.  
CC DOMAIN 411 416  
CC POLY-ASN.  
CC VARSPPLIC 102 107  
CC GLQPHL->VGNYIN (in isoform 2).  
CC  
CC FTFT 108 440  
CC /FTID=VSP\_007457.  
CC Missing (in isoform 2).  
CC FTFT 81 81  
CC R -> K (IN REF 1)  
CC  
CC CONFLICT 81 81  
CC

185 185 A -> T (IN REF. 1).  
193 193 A -> T (IN REF. 1).  
239 239 T -> S (IN REF. 1).  
344 344 V -> A (IN REF. 1).  
355 355 C -> R (IN REF. 1).  
394 394 TSVTN -> SSDTTT (IN REF. 1).  
406 406 H -> R (IN REF. 1).  
425 425 K -> N (IN REF. 1).  
440 440 K -> KVIK (IN REF. 1).  
SQ SEQUENCE 440 AA; 50623 MW; 0E0AB2C55517BE6F CRC64;  
Alignment Scores:  
Pred. No.: 1.14e-07 Length: 440  
Score: 217.50 Matches: 85  
Percent Similarity: 37.66% Conservative: 66  
Best Local Similarity: 21.20% Mismatches: 148  
Query Match: 2.83% Indels: 102  
DB: 1 Gaps: 12  
US-09-874-162a-7 (1-4409) x VRN2\_ARATH (1-440)  
QY 1399 AGAAGAAAGAGATCTCCAAATGAAACCCGACAAAATTAAGAAATATTTTATCAGTTT 1458  
Db 58 ArgLysArgLys-----SerArgSerThrGlyMetValValPheAsnThr 72  
QY 1459 CTCTATAACAACATACAGGCAACAACTGACGACGAGATGACCTGCATTGCCCTTGG 1518  
Db 73 LysAspCysAsnAsnThrLeuGlnArgThrGluValArgLysCysSerCysProPhe 92  
QY 1519 TGTACTCTGAAGTCCGCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCATAGC 1578  
Db 93 CysSerMetLeuCysGlySerPheLysGlyLeuGlnPheHisLeuAsnSerSerHisAsp 112  
QY 1579 AGATTATCTTC----- 1590  
Db 113 LeuPheGluPheGluPheLysLeuLeuGluGluThrValAsnValSerValLys 132  
QY 1591 -----AATGATGTTTATCATCCAAAGGTGTAGGATAGATGTT----- 1629  
Db 133 LeuAsnSerPheIlePheGluGluGluGlySerAspAspLysPheGluProPheSer 152  
QY 1630 -----TCTATC 1635  
Db 153 LeuCysSerLysProArgLysArgArgGlnArgGlyGlyArgGlnAsnThrArgArgLeu 172  
QY 1636 AATGAGTGTAT----- 1647  
Db 173 LysValCysPheLeuProLeuAspSerProSerLeuAlaAsnGlyThrGluAsnGlyIle 192  
QY 1648 -----GATGGCTCTTATGCA---GGAAATCTCTCAGGATATTTCATGCCAACCT 1692  
Db 193 AlaLeuLeuAsnAspGlyAsnArgGlyLeuGlyTyProGluAlaThrGluLeuAlaGly 212  
QY 1693 GGATTTGCTTTTGTAGTCGCAACGAGCAGTTAAGAGAACACCTTATCAGCAT----- 1743  
Db 213 GlnPheGluMetThrSerAsnIlePro-----ProAlaIleAlaHisSerSerLeu 229  
QY 1744 -----ATTCCTGTGTCAGGCCCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1788  
Db 230 AspAlaGlyAlaLysValIleLeuThrThrGluAlaValValProAlaThrLysThrArg 249  
QY 1789 TTTCTTGAATCTGAAGATGGGAGAGTAGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1848  
Db 250 LysLeuSerAlaGluArgSerGluAla-----ArgSerHisLeuLeuLeuGlnLys 266  
QY 1849 CGTCTGTATTTCATAGTGATACCTGCTTACCTCTCTCCGTCACCAAGCAAGCAAGCAAG 1902  
Db 267 ArgGlnPheTyThrHisSerHisArgValGlnProMetAlaLeuGluGlnValMetSerAsp 286  
QY 1903 GTAGATAGTGAAGATGAAGAGATCCTGAATGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1962  
Db 287 ArgAspSerGluAspGluValAspAspValAlaAspPheGluAspArgGlnMetLeu 306







QY	1789	TTTCTTGAATCTGAAGATGGGAAGTAGAACAGCAAGAAACATATAGTAGTGGCCACAAT	1848
Db	526	GlulLeuHsAlaGluArgCysGluAlaLysArgLeuGluArgLeuLys	542
QY	1849	CGTCTGTATTTCCATAGTAGATACCTGCTTACCTCTCCGTCACCAAGAAATG	1902
Db	543	ArgGlnPheThrHisSerGlnThrMetGlnProMetThrPheGluGlnValMetSerAsn	562
QY	1903	GTGATAGTGAAGATGAAGAAAGATCCCTGAATGGCTTAAGACAGAAAAACCATTACACAAAT	1962
Db	563	GluAspSerGluAsnGluThrAspAspTyrAlaLeuAspLysSerGluArgLeuArgLeu	582
QY	1963	GAAGAGTTTCTCGATGTTAATGAAGGAGAGAGAAAGATGATGAACCTCTGGAATCTCCAT	2022
Db	583	GluArgLeuValGlyValSerGlyGluGluLysArgTyrMetTyrLeuTrpAsnIlePhe	602
QY	2023	GTCATGAAGCATGGGTTTATTGCTGACAAATCAAAATGAATCATGCTCTATGCTGTGTGTA	2082
Db	603	ValArgLysGlnArgValIleAlaAspGlyHisValProTrpAlaCysGluGluPheAla	622
QY	2083	GAA	2106
Db	623	LysLeuHisLysGluGluMetLysAsnSerSerSerPheAspTrpTrpArgMetPhe	642
QY	2107	AAGAAGAATTTATGTCGAACTTCATGCTTCATCTAGTCAGCATGCATGCTTAATCTT	2166
Db	643	ArgIleLysLeuTrpAsnAsnGlyLeuIleCysAlaLysThrPheHisLysCysThrThr	662
QY	2167	ATTAGCATATGTCAAATAGATAAAGCTGTTACCAAGCTCCGTAATGCAGCAAAATTA	2226
Db	663	IleLeuLeuSerAsnSerAspGluAla	674
QY	2227	GAAAAGGGGNAATCGTCTCCCTGCGAAACGAGAANAATACTGAA	2271
Db	675	ThrSerGlySerAlaAlaAsnAlaAsnAsnGlnGlnSerMetGlu	689

RESULT 9

GIN4 YEAST

ID	GIN4 YEAST	STANDARD;	PRY;	1142 AA.
AC	Q12253;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	30-NOV-2000 (Rel. 39, Last annotation update)			
DE	Serine/threonine-protein kinase GIN4 (EC 2.7.1.1-).			
GN	GIN4 OR YDR507C OR D9719.13.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
NC	NCBI_TaxID=4932;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99030835; PubMed=9813093;			
RA	Longline M.S., Fares H., Pringle J.R.;			
RT	"Role of the yeast Gln4p protein kinase in septin assembly and the			
RT	relationship between septin assembly and septin function.";			
RN	J. Cell Biol. 143:719-736 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,			
RA	Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,			
RA	Hunnicke-Smith S., Hymann R., Komp C., Lashkari D., Lew H., Lin D.,			
RA	Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,			
RA	Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,			
RA	Winant A., Yelton M., Bostein D., Davis R.W.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: May play a role in septin assembly.			
CC	-1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. NIM1			
CC	subfamily.			

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use.  
 CC -----

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

-----

CC EMBL; U33140; AAA75513.1; -;  
CC DR EMBL; U33057; AAB64949.1; -;  
CC DR PIR; S59359; S59359.  
CC DR HSSP; Q63450; IA06.  
CC DR GerMOnline; I40999; -.  
CC DR SGD; S0002915; GIN4.  
CC DR GO; GO:0005935; C:bud neck; IDA.  
CC DR GO; GO:0007117; P:bud growth; IGI.  
CC DR GO; GO:000135; P:septin checkpoint; IGI.  
CC DR GO; GO:000921; P:septin ring assembly; IGI.  
CC DR InterPro; IPRO00719; Prot\_kinase.  
CC DR InterPro; IPRO08271; Ser\_thr\_pkin\_AS.  
CC DR InterPro; IPRO02290; Ser\_thr\_pkinase.  
CC DR InterPro; IPRO01245; Tyr\_pkinase.  
CC DR Pfam; PF00069; pkinase\_1.  
CC DR PRINTS; PR00109; TYRKINASE  
CC DR ProDom; PD000001; Prot\_kinase; 1.  
CC DR SMART; SMC0220; S\_TKC; 1.  
CC DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
CC DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
CC DR PROSITE; PSS0011; PROTEIN KINASE DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 19 289 PROTEIN KINASE  
FT NP\_BIND 25 33 ATP (BY SIMILARITY).  
FT BINDING 48 48 ATP (BY SIMILARITY).  
FT ACT\_SITE 156 156 BY SIMILARITY.  
SQ SEQUENCE 1142 AA; 129857 MW; EC16PF4BB49DD811 CRC64;

Alignment Scores:

Pred. No.:	9 93e-05	Length:	1142
Score:	176.50	Matches:	152
Percent Similarity:	34.75%	Conservative:	102
Best Local Similarity:	20.79%	Mismatches:	264
Query Match:	2.30%	Indels:	213
DB:	1	Gaps:	35

US-09-874-162A-7 (1-4409) x GINA\_YEAST (1-1142)

QY	172	GATCACGGGTTTGAAGAAACAAGAATTACAGACCA-	-----ACC 213
Db <td>337</td> <td>AspProGluGlutyle--LysGluLysLeuArgGluProGlyAlaAsnAlaGluLysThr</td> <td>355</td>	337	AspProGluGlutyle--LysGluLysLeuArgGluProGlyAlaAsnAlaGluLysThr	355
QY <td>214</td> <td>TARGTGCGCTGAGTTACATAAATAGATTTCATCACAGATGCTGCCCGCCGAGACGAGGAG</td> <td>273</td>	214	TARGTGCGCTGAGTTACATAAATAGATTTCATCACAGATGCTGCCCGCCGAGACGAGGAG	273
Db <td>356</td> <td>LeuTyraLeuLeuTyr-----ArgPheLysCysAspThr-<td>369</td></td>	356	LeuTyraLeuLeuTyr-----ArgPheLysCysAspThr- <td>369</td>	369
QY <td>274</td> <td>TCOCTTAAGAAGAAGATTACGCCGAGACTCTGCCTGACTCTGTCCAAGC---<td>330</td></td>	274	TCOCTTAAGAAGAAGATTACGCCGAGACTCTGCCTGACTCTGTCCAAGC--- <td>330</td>	330
Db <td>370</td> <td>GluLeuIleLysglnglInValLysLysArgGlnSerIleSerValSerValSer</td> <td>389</td>	370	GluLeuIleLysglnglInValLysLysArgGlnSerIleSerValSerValSer	389
QY <td>331</td> <td>CGAGGG---AANTGTGTCGACTCCCCCA-<td>372</td></td>	331	CGAGGG---AANTGTGTCGACTCCCCCA- <td>372</td>	372
Db <td>390</td> <td>ProSerLysLysValSerThrThrProGlnArgArgAsnArgGluSerLeuIleSer</td> <td>409</td>	390	ProSerLysLysValSerThrThrProGlnArgArgAsnArgGluSerLeuIleSer	409
QY <td>373</td> <td>CCTTAGCTCCCCTGGTACCACCACCCCATC-----ACCCCTCTCTCT</td> <td>411</td>	373	CCTTAGCTCCCCTGGTACCACCACCCCATC-----ACCCCTCTCTCT	411
Db <td>410</td> <td>ValThrSerSerArgLysLysProIleSerPheAsnLysPheThrAlaSerSerAlaSer</td> <td>429</td>	410	ValThrSerSerArgLysLysProIleSerPheAsnLysPheThrAlaSerSerAlaSer	429
QY <td>412</td> <td>TCATTCGCGAGCAGCACTCCGACAGACCAACACAGATCTAGATTTCTCGAACCTCGG</td> <td>471</td>	412	TCATTCGCGAGCAGCACTCCGACAGACCAACACAGATCTAGATTTCTCGAACCTCGG	471
Db <td>430</td> <td>SerSerAsnLeuThrThrProGlySerSerLysArgLeuSerLysAsnPheSerLys</td> <td>449</td>	430	SerSerAsnLeuThrThrProGlySerSerLysArgLeuSerLysAsnPheSerLys	449
QY <td>472</td> <td>AATCTCATAGACCAATAATTTTTGCACAGAACTCTTACTTACATGCTCATCGAACTCC</td> <td>531</td>	472	AATCTCATAGACCAATAATTTTTGCACAGAACTCTTACTTACATGCTCATCGAACTCC	531
Db <td>450</td> <td>LysLysLeuSerThrlleValasnGlnSerProThrProalaser---ArgAsnLys</td> <td>468</td>	450	LysLysLeuSerThrlleValasnGlnSerProThrProalaser---ArgAsnLys	468
QY <td>532</td> <td>AGAACCAACATC-----AAAAAGAAAACATTAAAGTTTGATGATGATTATCAAAGTA</td> <td>585</td>	532	AGAACCAACATC-----AAAAAGAAAACATTAAAGTTTGATGATGATTATCAAAGTA	585

Db 469 ArgAlaSerValIleAsnValGluLysAsnGlnLysArgAlaSerIlePheSerThrThr 488  
 QY 586 GAGAAATGAAGAGAGAGCAAGATCT----- 612  
 Db 489 LysLysAsnLysArgSerArgSerIleLysArgMetSerLeuIleProSerMetLys 508  
 QY 613 ---CATAGCTGTGCTCATTCCAGCTTACGTTTACTGTTTCTTCCACAAAATGAT 669  
 Db 509 ArgGluSerValThrThrLysLeuMetSerThrThrLysLeuAlaGluAspAsp 528  
 QY 670 AAGCCATCACCACAACTCAGAAATGAACAAATCTCTTACCCCTGGAGTCTGCTGTG 729  
 Db 529 TrpGluTrpIleGluLysGluThrLysArgThrSerSerAsnPheAlaThrLeuIleAsp 548  
 QY 730 AAAGTT-----TGCCACAAAACAAAGAGATGAAGTTGT 765  
 Db 549 GluIlePheGluTrpGluLysTyGluGlnIleArgLysGluLysGluGluLeuArg 568  
 QY 766 CCAATAAGGCAAGTCCACAGGT----- 789  
 Db 569 LysValArgGluAlaLysAlaArgGluGluLysGluArgArgLysGlnGlu 588  
 QY 789 ----- 789  
 Db 589 LysGluArgAlaArgLysLeuLeuGluLysGluAspLeuLysArgLysGlnGluLeu 608  
 QY 790 AAAAAGCAGGTGCTTTG---ATTCTGACCTCAATCAA----- 825  
 Db 609 LysLysGlnIleGluIleAspIleSerAspLeuGluGlnGluLeuSerLysHisLysGlu 628  
 QY 826 ---ACAAAACCCGGAATTTCCCTCCCTGCGATTTCCAGTAAGTAATTTGAACCTAGT 882  
 Db 629 GluLysLeuAspGlyAsnIleArgSerIleSerAlaProMetGluAsnGluLysAsn 648  
 QY 883 AACGCCATATGGTGAAGTCTTACTCGTTGCTATTAGA-----GTG 924  
 Db 649 IleAsnHisLeuGluValAlaPheAspAsnIleLeuArgArgArgAsnPheSerLeuGln 668  
 QY 925 ACTGCTCCAGGAAGAAGAGAGTTTAATGGAATGATTAATGAGAAACCAATGAATATT 984  
 Db 669 ThrArgProValSerArgLysAspProGlyIleMetPheSerSerProThrGluGluVal 688  
 QY 985 GATGTCAATGAAGACTTCCAGCAGAGAAACAAATCCTGAGGATGGGAAAGACA 1044  
 Db 689 SerProValGlu-----ProLysArgThrGluAsnGluArgLeuThrThrGluLys 706  
 QY 1045 TTGTGTGCACAAATG-----ACAGTATTGTGATAAAACAGCGCG 1083  
 Db 707 IleLeuGluThrIleArgArgSerLysPheLeuGlySerSerPheAsnIleAspLysGlu 726  
 QY 1084 TTACAGCTTTAGTGGGATATGAA-----GTAGCCATGCGAGGAATG---GAGAA 1134  
 Db 727 LeuLysLeuSerLysMetGluTrpProSerIleIleAlaProGlnArgLeuSerGluGlu 746  
 QY 1135 TGTCCAATAAGCAAGAAAGACACATGGAGACTATTCT-----GATGGGAAG 1185  
 Db 747 ArgValValSerAspSerAspGlyTyroGluSerIleLeuProLysAspGlyAsn 766  
 QY 1186 AGGCTGCTCCATTCGAA-----ACATTTTCTCAGGACCTACGTTGCGAGTTCACCTCTT 1239  
 Db 767 GlyValSerGlnLeuLysAspSerThrAlaThrThrAlaProValSerAspGlyArgLeu 786  
 QY 1240 GTTGGCAGGAGACCAATGATAAATCTACGGCTCCTATTGCAACCTCTTGGCCACT 1299  
 Db 787 ArgLysIleSerGlu-----IleArgValProGlnPheThr 798  
 QY 1300 AGAAATTCAGAGAGTCCATCAGGAAACAAAGCTGTTGCTGATTAACCTTACTCAAACT 1359  
 Db 799 ArgLysSerArgHisPheSerGluSerAsnLysArgLeuSerVal-----LeuSerMet 816  
 QY 1360 ATTGCTGTTAAGATCATCTACATGATCTACAAACAAAGAAAGGATACTCCA 1419  
 Db 817 TyrSerThrLysGluSerPheThrAsnLeuValAspIleLeuLysAsnGlyAsnLeuAsp 836

QY 1420 AATGAAACCGCAACAAATTAAGAATATTTTTATCAGTTTCTCTATAACAACAATACAAG 1479  
 Db 837 ValAsnAsnGlnGlnSerGlnArgIle----- 845  
 QY 1480 CAACAAACTGAAGCAAGAGATGACCTGCATTCGCCCTTGGTGTACTCTGAACTCGCGCAA 1539  
 Db 846 ProThrProArgSerAlaAspAsp----- 853  
 QY 1540 CTTTATAGTTTACTCAACCATCTTAACTCTGCCATAGACAGATTTATCTTCAACTATGTT 1599  
 Db 854 -----SerGluPheLeuPheGlu----- 859  
 QY 1600 TATCATCAAAAGGTGCTTAGGATAGATGTTTCTATCAATAGTGTATGATGGCTCTCTAT 1659  
 Db 860 -----ThrValAsnGlu-----GluAlaGluTrp 867  
 QY 1660 GCAGAAATCCTCAGGATATTTCATCGC----- 1686  
 Db 868 ThrGlyAsnSerSerAsnAspGluArgLeuTyrAspValGlyAspSerThrIleLysAsp 887  
 QY 1687 -----CAACCTGGATTTCCT-----TTTACTCGCAACGCGACCACTTAAGAGA 1728  
 Db 888 LysSerAlaLeuLysLeuAsnPheAlaAspArgPheAsnGlySerAsnGluAlaLysGln 907  
 QY 1729 ACACCTATCACATATTTCTGTGTGCGAGGCCAAACGAAACAAAGCAGCATGCTCAAA 1788  
 Db 908 ThrAspAsnLeuHisLeuProIleLeuProLeuAsnGlyAspAsnGluLeuArgLys 927  
 QY 1789 TTTCTTGAATCTGAAGATGGGAGTAGAAGCAAGCAACAAACATATAGT----- 1836  
 Db 928 -----GlnAsnSerGlnGluGlyAspGlnAlaHisProLysIleLysSerMetIleProGlu 946  
 QY 1837 AGTGCCCAACATCGTCTGTATTTCATAGTAGATACCTGCTTACCTCTCCGTCACCAAGAA 1896  
 Db 947 SerGlySerSerSer-----HisThrGlu-----LysGlu 956  
 QY 1897 ATGGAAGTAGATAGTAGAAGATGAAGAGATCTGTAATGG----- 1935  
 Db 957 GluGluAsnGluGluGlyGluLysProGluGlnHisLysGlnGluLysAspGln 976  
 QY 1936 ---CTAAGAGAAACCAACCATTAACAAATGAA 1965  
 Db 977 GluLysArgGluLysValValAspAspMetGlu 987  
 RESULT 10  
 Y373 BOVIN  
 ID Y373 BOVIN STANDARD; PRT; 1453 AA.  
 AC Q9TU23;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein KIAA0373.  
 GN KIAA0373.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 CX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Trachea;  
 RA Jovov B., Ripoll P.J., Benos D.J.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

DR EMBL; AF176816; AAF00990.1; -.  
KW Hypothetical protein; Coiled coil.  
FT DOMAIN 37 1426 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF4139 CRC64;

Alignment Scores:	
Pred. No.:	0.000716
Score:	164.50
Percent Similarity:	35.15%
Best Local Similarity:	19.73%
Query Match:	2.14%
DB:	1
Length:	1453
Matches:	174
Conservative:	136
Mismatches:	333
Indels:	239
Gaps:	39

US-09-874-162A-7 (1-4409) x Y373 BOVIN (1-1453)

QY	199	TTACAGCAGCCAACTTATGTTCCCTCGAGTTACATAAATAGATTTCATGCAGATGTGTGCC	258
Db	628	LeuAsnGlnLysGlueLuValLeuLysIstyr-----GlnHisLeuLeuGluLysAla	645
QY	259	CGCGCAGACGACGAGTCCCTTAAGAAGAG--:::---ATCAGCG	297
Db	646	ArgGluGluGlnArgGluLeuValLysIstHsGluGluLeuHisThrLeuHisArg	665
QY	298	AAGCTTCGCTGACTGTGTCAGCTCAGTGTCGAGGGGAATGTGTCCACTCCGCCACGC	357
Db	666	LysLeuGluLeuGlnAlaAspAsnSerLeuSerLysPhe:::     :::	679
QY	358	CACAGCAGTGGGAAGCTTACTCCCCCGTAGCCACCACCCCTCCCTCTTCAATC	417
Db	680	GluThrAlaTrpAspLeuLeuLysGlnSerProThrProValProThrAsnLysHisPhe	699
QY	418	CGCAGCAGCCTCCGACGAGCCAACACAGATCTATAGATTTCCTCGAACTCGGAATCTC	477
Db	700	IleArgLeuAlaGluMetGlnThrValAlaGluGlnAspSerLeuSerLeu	719
QY	478	ATAGCACCAATATTTTGCACAGAACTCTTACTTACATGTCTCATCGAACTCCAGAAC	537
Db	720	Val-----IleLysLeuLysGlnValSerGlnAspLeuGluArgGlnLysGluLeuThr	737
QY	538	AACATCAAAGGAAAACATTAA-----GTTGATGATATGTTATCA	579
Db	738	GluLeuLysIleLysGluPheGluAsnMetLysLeuArgLeuGlnGluAsnHisAlaAap	757
QY	580	AAAGTAGAGAAAAATCAAGGAGAG-----:::     :::	603
Db	758	GluValLysLysIleLysAlaGluValGluAspLeuArgCysLeuLeuValGlnSerGln	777
QY	604	CAAGATCTCATAGTGTGTCACTCATTTGCAG-----:::     :::	636
Db	778	LysGluSerGlnSerLeuLysSerGluLeuGlnThrGlnLysGluAlaAsnSerArgAla	797
QY	637	-----CTTACGTTTTACTGGTTCTTCACAAAAATGATPAAGCCATCCAAACTCAGAA	690
Db	798	ProThrThrMetArgAsnLeuValGluArgLeuLysSerGlnLeuAlaLeuLysGlu	817
QY	691	AATGAACAAATTCGTTACCTCGAAGTCCTGCTGTGAAA-----:::     :::	732
Db	818	LysGlnGlnLysAlaLeuSerArgAlaLeuLeuGluLeuArgAlaGluMetThrAlaAla	837
QY	733	-----GTTTGGCCACAAAAAAGAGAGATGTAAGTTGTCATPAAGCAA	777
Db	838	AlaGluGluArgIleIleSerMetThrSerGlnLysGluAlaAsnLeuAsnValGlnGln	857
QY	778	GTTCCCACA-----GGTAAAGACAGGTGCTTTGATCTCTCACTCAATCAA-----	825
Db	858	IleValAspArgHisThrLysGluLysSerGlnIleGluAspLeuAsnGluAsnIle	877
QY	826	-----ACAAAAACCCGGAATTTCCCGTCCCTTCGATTCCTCC	861
Db	878	LeuLysLeuLysGluAlaLeuLysThrSerLysAsnArgGluAsnThrLeuThrAspAsn	897
QY	862	AGTAATGAATTTGAACCTAGTAAACGCCATATGTTACTCGTTGCTATTTAGA	921

898	LeuAsnAspLeuThrAsnGluLeuGlnAsnLysGlnLysAlaTyrGlyLysValLeuArg	917
922	GTCACTCGTCCAGGAAGAAGAGAGTTTAATGCAATGATTAATGAGAAACCAATGAAAT	981
918	-----GluLysAspAlaVal	922
982	ATTGATGTCATGAAGACCTTCCAGCCAGACAGAAAGAAACAAATCCTGAG-	1039
923	AspGlnGluAsnAsnGluLeuLysArgGlnIleLysArgLeuThrSerGlyLeuGlnGly	942
1030	-----GATGGCGGAAAAGACATTTCTGCACAAATGCACAGTATTGTATAAAAC	1077
943	LysProLeuIleAspAsnLysGlnSerLeuIleGluLeu-	957
1078	AGCGCCTTACAGCTTTTAGATGGGAAATATGAAGTAGCCATGCAGGAAATCGAAGATGT	1137
958	LysLysIleLysLysLeuGluSerGlnLeuGluArgLysValAspGluAlaGluMetLys	977
1138	CCAATAAACAAGAAAGAGCA-	1164
978	PrometLysGluLysSerAlaArgGluGluValIleArgTrpGluGluGlyLysLysTrp	997
1165	GAGACTATTCTTGATGGGAAGAGCGCTCCATTCGAAACATTTCTCAGGAGCACTACG	1224
998	GlnThrLysIleGluGlyIleArgAsnLysLeuLysGlu-	1015
1225	TTCAGTTCACCTCTCGTTGCAGCAGGAGACCAATGATAAATCTACGGCTCCTATTGCC	1284
1016	TyrIleLeuThr-	1026
1285	AAACCTCTTGCCACTAGAAATTACAGAGAGTCTCCATCAGGAAACCAAGCCTGGTTCAGTT	1344
1027	AspLeuPheAlaLysAlaAspLysGluLysLeuThrLeuGlnArgLys-	1043
1345	AAACCTACTCAACTATTGCTGTTAAAGATCATTGACTACAGAT-	1389
1044	LysThrThr-	1056
1390	-----CTACAAACAGAAAGAAAGGATCTCCAAATGAAACCCGACAAAAATTAAGA	1443
1057	ArgValLeuGluSerGluLysGluLeuGluLysLysArgAsnLeuAspLeuGlu	1076
1444	ATATTTCATGTTCTTCTATACACAATATACAGGCAACAATGAAGCAAGAGATGAC	1503
1077	AsnAspIleSerTyrMetArgSerHisGlnAlaLeuProArgAspSerValIleGluAsp	1096
1504	CTGATTCGCCCTGGTCTACTCTGAACCTCGCGCAACTTATAGTTTACTCAAGCATCTT	1563
1097	LeuHisLeuGlnAsnLysTyrLeuGln-	1115
1564	---AAACTCTGCCATACGATTTATCTTCAACTATGTTTATCATCCAAAGGTGCTAGG	1620
1116	SerLysAspAlaTyrSerArg-	1127
1621	ATACATGTTCTATCAATGATGTATGATGGCTCTCTATCCAGGAATCTCAGGATATT	1680
1128	IleAsp-	1140
1681	CATGCCAACCTCGATTTCGTTTTAGTCGCAACGGACCAAGTTAAG--AGAACACCTATC	1737
1141	GlnArgGluAsnLeuLysLeuSerSerGluAsnIleGluLeuLysPheGlnLeuGluGln	1160
1738	ACACATATTCTGTGTGCAGGCCAAA-	1788
1161	AlaAsnLysAspLeuProArgLeuLysAsnGlnValArgAspLeuLysGluMetCysGlu	1180
1789	TTTCTTCAATCTCAAGATGGGGAAGTAGACACCAA-	1842
1181	PheLeuLysLysGluLysAlaGluValGluArgLysLeuGlyArgValargGlySerGly	1200
1843	CACATCGCTGTATTCTCCACTAGTANACTGCTTACCTCCGTCCTCCACAAGAAATGGA	1902
1201	ArgSer-	1202



```

Db      316 luLeuAsnSerIleArgGluLeuAsnThrAlaLys-----ValIleAlaA 331
QY      857 TTTTCAGTAATGAATTTGAACCTAGTAACAGCCATATGGTGGAAGCTTACTCGTGCTAT 916
Db      331 spAspSerLysGlnThrProGluAsnGluAspLeuLeuGlnLeuThrL 351
QY      917 TTAGAGTGAAGTCCAGGAGAGAGAGAGATTTAATGGAATGATTAAATGAGAAACCAATG 976
Db      351 ysGluLysLeuAlaGlnCysGluLysGluCysLeu---ArgLeuSerSerIleThrAspG 370
QY      977 AAAATATTGATGTCATGAAGAGCTTCAGCC----- 1008
Db      370 luAlaAspGluAspAsnGluAsnLeuSerAlaLysSerSerSerAspPheIlePheLeuL 390
QY      1009 -----AGAGAAACAGAAATCGTGAGATGGGGAAGAAGACAT 1045
Db      390 ysLysGlnLeuIleLysGluArgArgThrLysGluHisLeuGlnAsnGlnIleGluThrP 410
QY      1046 TTGTTGCACAAATG-----ACGATATTGATAAAMAAACAGCGCTTACAGC 1090
Db      410 heIleValGluLeuGluHisLysValProIleIleAsnSerPheLysGluArgThrAspM 430
QY      1091 TTTTAGATGGGAATATGAATGAGCATGCAGGAATGGAAGATGTCCAATAAGCAAGA 1150
Db      430 etLeuGluAsnGluLeuAsnAsnAlaAlaLeuLeuGluHisThrSerAsnGluLysA 450
QY      1151 AAAGAGCAACATGGAGACTATTTCTTGATGGGAAGAGCGCTCCATTCGAAACATTTT 1210
Db      450 snAlaLysValLysGluLeuAsnAlaLysAsnGlnLysLeuValGluCysGluAsnAspL 470
QY      1211 CTCAGGACCTACG-----TTTCAGTTCTACTCTTCGTT 1243
Db      470 euGlnThrLeuThrLysGlnArgLeuAspLeuCysArgGlnIleGlnThrLeuLeuIleT 490
QY      1244 GGACAGGAGACCAATGATAATCTACGGTCTCTATTGCCAAA----- 1287
Db      490 hrAsnSerValSerAsnAsp---SerLysGlyProLeuArgLysGluGluIleGlnPheI 509
QY      1287 ----- 1287
Db      509 leGlnAsnIleMetGlnGluAspAspSerThrIleThrGluSerAspSerGlnLysValV 529
QY      1288 -----CCCTTGCCCACTAGAAATTC----- 1308
Db      529 alThrGluArgLeuValGluPheLysAsnIleIleGlnLeuGlnGluLysAsnAlaGluL 549
QY      1309 -----GAGAGTCTCCATCAGGAAACCAAGC 1333
Db      549 euLeuLysValValArgAsnLeuAlaAspLysLeuGluSerLysGluLysSerLysG 569
QY      1334 CTGGTTTCAGTTAAA---CCTACTCAACTATT---GCTGTTAAAGAAATCATTTGACTACA- 1386
Db      569 lnSerLeuGlnLysIleGluSerGluThrValAsnGluAlaLysGluAlaIleIleThrL 589
QY      1387 -----GATCTACAAACAAGAAAGAAAG----- 1410
Db      589 euLysSerGluLysMetAspLeuGluSerArgIleGluLeuGlnLysGluLeuGluG 609
QY      1411 -----GATACTCAATGAACCCAGCAAAATTAAGAAATATTTATCAGTTTC 1459
Db      609 luLeuLysThrSerValProAsnGluAspAlaSer----- 620
QY      1460 TCTATAACACAAATACA---AGGCCAACAACTGAAGCAAGAGATGACCTGATTCGCCCTT 1516
Db      621 ---TyrSerAsnValThrIleLysGlnLeuThrGluThrLysArgAspLeuGluSerGlnV 640
QY      1517 GGTGACTCTGAAC-----TGCCGCAAACTTTATAGTTTACTCAAGC 1558
Db      640 alGlnAspLeuGlnThrArgIleSerGlnIleThrArgGluSerThrGluAsnMet-Ser 659
QY      1559 ATCTTAACCTGCGCATAGAGATTATCTTCAACTATGTTTATCATCCAAAGCGTCTA 1618

```

```

Db      660 LeuLeuAsnLysGluIleGlnAspLeuTyrAspSerLysSerAspIleSerIleLysLeu 679
QY      1619 GGATAGATGTTTCTATCAATGAGTGTTATGATGCTCTCTATGATGAGAAATCTCTCAGATA 1678
Db      680 Gly-----LysGluLysSerSerArgIleLeuAla-----GluGluArgPheLysLeu 695
QY      1679 TTCATGCCCAACCTGGATTGCTTTAGTCGCAACGACGACGATTAAGAGAACACCTATCA 1738
Db      696 LeuSerAsnThrLeuAspLeuThrLysAlaGluAsnAspGlnLeuArgLysArgPheAsp 715
QY      1739 CACATATT-----CTTGTCGTCAGCGCCAAACGAAACAAACAAACAGCATGCTGAATTT 1791
Db      716 TyrLeuGlnAsnThrIleLeuLysGlnAspSerLysThrHisGluThrLeuAsnGluTyr 735
QY      1792 CTT-----GAATCT 1800
Db      736 ValSerCysLysSerLysLeuSerIleValGluThrGluLeuLeuAsnLeuLysGluGlu 755
QY      1801 GAGATGCGGGAAGTAGAACAGCAACACATATAGTAGTGGCCACAAATCGTCTGTATTTTC 1860
Db      756 GlnLysLeuArgValHisLeuGluLysAsnLeuLysGlnGluLeuAsnLysLeuSerPro 775
QY      1861 CATAGTAGTACCTGCTTACCTCTCCGTCACAGAAATGGAAGTAGATAGTAGAAGTAA 1920
Db      776 GluLysAspSerLeuArgIleMetValThrGlnLeuGlnThrLeuGlnLysGluArgGlu 795
QY      1921 AAGATCTCTGAATGGCTAAAGAGAAACCAATTCACAAATTCGACAGATTTTCTGAT--- 1977
Db      796 AspLeuLeuGluGluThrArgLysSerCysGlnLysLysIleAspGluLeuGluAspAla 815
QY      1978 GTTAATCAGGAGAGAGAGAGAGTGAACACTCGAAATCTCCATGTCATGATGAGCATGG 2037
Db      816 LeuSerGluLeuLysLysGluThrSerGln-----LysAspHisHisLysGln--- 832
QY      2038 TTTATTGCTGACAAATCAATGAATCATGCTGCTGATGCTGTTGTGAGAAATATATGACAG 2097
Db      833 LeuGluGluAspAsnAsnSerAsn-----IleGluThrPyrGlnAsn 846
QY      2098 AAAATA-----ATTGAAGAAATTTATGTCGAAACTTCATGCTTCATCTAGTCAGCATG 2151
Db      847 LysIleGluAlaLeuLysLys----- 853
QY      2152 CATGACTTTAATCTTTATTAGCATAATGTCATAGATAAAAGCTGTTACCAAGCTCCGTGAA 2211
Db      854 ---AspTyrGluSerVal---IleThrSerValAspSerLysGlnThrAspIleGluLys 871
QY      2212 ATCAGCAAAATAGAAAGGGGGAATCTGCTCCCTGCCAAACGAAAGAAATTAAGTAA 2271
Db      872 LeuGlnTyrLysValLysSerLeuGlu-----LysGluIleGluGlu 885
QY      2272 GAACAAATGGGACACAAATGATTTAGTGAATTAACCTCAAAAGAGAAAGCTTTGGAA 2331
Db      886 AspLys-----IleArgLeuHisThrTyrAsnValMetAspGluThrIleAsn 901
QY      2332 ACAGATAGTCTCTCAGGGGTTTCAAAACAGACAGCAAAACCAAACTCTGAAAAGCTCTA 2391
Db      902 AspAspSer-----LeuArgLysGluLeuGluLysSerLysIle-----AsnLeu 916
QY      2392 ACCCATGTTATGGACAA 2409
Db      917 ThrAspAlaTyrSerGln 922

```

RESULT 12

UBP3 YEAST

ID\_UBP3 YEAST

AC Q01477;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin

DE thiolesterase 3) (Ubiquitin-specific processing protease 3)

DE (Deubiquitinating enzyme 3).

GN UB3 OR YER151C.







Db	567	ysThrLeuSerThrIleProLysPheLysAspLeuGlnTrpGlyHisGlnGluAspAlaG	587
QY	2006	AACTCTGGAAATCTCCATGTCATGAAGCAT	2038
Db	587	lucGluPheLeuThrHisLeuLeuAspGlnLeuHisGluGluLeuIleSerAlaIleAspG	607
QY	2039	TTATTCTGCATCAATCAATGAATCATGCCTGTATGCTGTTTGTAGAAAAATTACGACAGA	2098
Db	607	lyLeuThrAspAsnGlu	619
QY	2099	AAATAATTAAAGAAATTTATGTCGAAACTTCATGCTTCATCTAGTCAGCATGCAT	2154
Db	619	erIleAsnAspGluGlnLeu--LysValPhePheIleArgAsnLeuSerArgTyrGlyL	638
QY	2155	----GACTTTAATCTTATTAGCATTAATGTCATAGATAAAGCTGTTACCAAGCTCCGTG	2209
Db	638	ysAlaGluPhe	651
QY	2210	AAATGCAGCAAAAATTAGMAAAGGGGAATCTGCTTCCCTCGCAACGAAAGAAATAACT	2269
Db	651	lueuIleGluLysTyr	665
QY	2270	AGAAACAAAATGGAGCAGCAAAATGGATTTAGTGAATTAATCTCAAAAGAAAGCTTTGG	2329
Db	665	lucGlu	671
QY	2330	AAACAGATAGTCTCAGGGGTTTCAAAACAGACAGCAAAAACAAAAA	2376
Db	672	-----ValSerGlySerLysArgLysGlyLysThrLys	683
RESULT 13			
PCP1 SCHPO			
AC	ID	PCP1 SCHPO	STANDARD; PRT; 1208 AA.
AC	Q92351		
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	15-MAR-2004	(Rel. 43, Last annotation update)	
DE	Spindle pole body protein pcpl1		
GN	PCP1 OR SPAC69.06C.		
OS	Schizosaccharomyces pombe (Fission yeast)		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
OX	NCBI_TaxID=4896;		
RN	[1]		
RN	SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.		
RX	MEDLINE=21852775; PubMed=11864908;		
RX	FLORY M.R., Morphew M., Joseph J.D., Means A.R., Davis T.N.;		
RT	"Pcp1p, a Spcl10p-related calmodulin target at the centrosome of the		
RL	fission yeast Schizosaccharomyces pombe.";		
RL	Cell Growth Differ. 13:47-58(2002).		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=972;		
RX	MEDLINE=21848401; PubMed=11859360;		
RX	WOOD V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,		
RA	SCOURAS J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,		
RA	BROOKS K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	COLLINS M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,		
RA	GENTLES S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	HOLROYD S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,		
RA	JAMES K., Jones L., Jones M., Leather S., McDonald S., McLean J.,		
RA	MOONEY P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,		
RA	OLIVER K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,		
RA	RUTHERFORD K., Rutter S., Saunders D., Seeger K., Sharp S.,		
RA	SKELTON J., Simmonds R., Squares R., Squares S., Stevens K.,		
RA	TAYLOR K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,		
RA	WOODWARD J., Volckaert G., Aert R., Robben J., Grymonprez B.,		
RA	WELTJENS I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,		
RA	GABEL C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,		
RA	BORZYM K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,		
RA	EGER P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.		

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
 RT "The genome sequence of *Schizosaccharomyces pombe*,"  
 RL Nature 415:871-880 (2002).  
 CC -!- FUNCTION: Spindle pole body component that binds calmodulin.  
 CC Overexpression of pcp1 causes the formation of supernumerary SPB-  
 CC like structures and disrupts both mitotic spindle assembly and  
 CC chromosome segregation.  
 CC -!- SUBCELLULAR LOCATION: Spindle pole body.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; Z81317; CAB03608.1; -;  
 DR EMBL; AF348506; AAK31344.1; -;  
 DR PIR; T39068; T39068.  
 DR GeneDB SPombe; SPAC609.06c; -;  
 KW Calmodulin-binding; Coiled coil.  
 FT DOMAIN 151 375 COILED COIL (POTENTIAL).  
 FT DOMAIN 387 803 COILED COIL (POTENTIAL).  
 FT DOMAIN 874 1091 COILED COIL (POTENTIAL).  
 FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;

## Alignment Scores:

Pred. No.:	0.00236	Length:	1208
Score:	157.00	Matches:	184
Percent Similarity:	35.51%	Conservative:	153
Best Local Similarity:	19.39%	Mismatches:	300
Query Match:	2.04%	Indels:	312
DB:	1	Gaps:	40

## US-09-874-162a-7 (1-4409) x PCPI\_SCHPO (1-1208)

QY	127	CTGGCCGACCTCATGACACATCGAGGACCAACACCATCGATACAGATCCACGGGTTTA	186
DB	127	LeuThrAspLeuSerLysAsnSerGluIleAspHisThrAspGly-----	141
QY	187	GAATAACAGAAATACAGCCCAACCTATGTTGCTGCTGATGATACATAATAGATTCTATG	246
DB	142	-----GluLeuProAlaAsnAlaAlaLeuThrLeu-----	151
QY	247	ACAGATGCTGCCGCGGACGAGAGTCCCTAAGAGAGAGATTTCAGCCGAGCTCTCG	306
DB	152	-----ArgGluGlnGluLysValLeuGluLysVal-SerArgGluAsnPh	166
QY	307	CTGACTC--TGTCAGCTAGTCTCTGAGGGAATGTGTCACCTCCCGCCACGACG	363
DB	166	eGlyLeuArgIleLysIleValCysLeuGlu-LysArgLeuGluSerMetAlaProGluG	186
QY	364	AGTGGAGCCCTTACTCCCGCGTACCCACCCATCCCTCTCTTCTTCCGCGAGC	423
DB	186	inLeLysGlu--AlaValLysAspAsnVal-----GluL	197
QY	424	AGCACTCCGACAGACCCACACAGATCTATAGATTCTTGAATCTCGAATCTCATGCA	483
DB	197	euHisAlaGluArgAlaAsnLeuGlnLeu-GlnLeuLysArgThrGluSerLeuLeu---	215
QY	484	CCAATATTTTGCACAGAACTCTTACTACATGCTCATCGAACTCCAGAACCAACATC	543
DB	216	-----GlnLysSer	218
QY	544	AAAGGAAACATTTAAAGTTGAT-----GATATGTTATCAAAAGTAGAGAAATG	594

DB	219	GluAspLysAsnPhLeuLysLeuGluLysValAspTyrLeuSerLysValAsnAspVal	238
QY	595	AAAGGAGACAGAAATCTCATAGCTTGTGACGCTCATTTGACGCTTACGTTTACGTTTC	654
DB	239	GluGlnSerGlnAsnValLysValPheThrGluArgIleArg-----	252
QY	655	TTCACAAAAAT-----GATAGCCATCCACAACTCAGAAAT	693
DB	253	PheLeuGluAsnAlaLeuGluLysValGlnArgGluLysAspSerLeuThrGluMet	272
QY	694	GAACAAATTTCTGTTACCTGGAAGTCCTGCTGTGAAAGTTTGCACAAAAAGAAAG	753
DB	273	GluGluAspLysSerAsnLysGluVal-----	281
QY	754	GATGTAAGTTGTCCATAAGGCAAGTTCCACAGGT-----AAAAAGCAGGTG	801
DB	282	AspTyrGluTyrGluIleArgGlnLeuGlnAsnArgLeuAspGluLeuSerGluGluLeu	301
QY	802	CCTTTGATTCTGACCTCAATCAACAAACCGGAAATTTCCCTGCTTCCGCTTCC	861
DB	302	AspValAlaGlnAspLeuThrGluLysGluAspGluIleAlaThrLeuLysArgGln	321
QY	862	AGTAATGAATTTGAACCTAGTAACAGC-----	888
DB	322	IleGluGluLysGluAsnSerSerAlaPheGluAsnGluGluAsnSerTyrVal	341
QY	889	CATATGGTGAAGTCTTACTGCTGTTGTTATTTAGAGTCTGCTCCAGGAGAGAGATTT	948
DB	342	HisLeuGlnGluAspTyrAlaIleLeu-----	350
QY	949	AATGGAATGATTAAATGGAGAACCAATGAAATATTTGATGTCATGAAGAGCTTCCAGCC	1008
DB	351	-----GlnAlaLysCysAspGluPheAlaAspArgIleGlnValLeuThrAla	366
QY	1009	AGAAGAAACGAAATCGTGAG-----	1029
DB	367	AspLeuGluLysGluLysGluAsnGlnIleMetHisGluSerGluAlaSerIleGlyLeu	386
QY	1030	---GATGGGAAAGACATTTGTCACAAATGACAGTATTTGATAAAACAGGGGCTTA	1086
DB	387	ThrAspSerMetGlnValHisThrLeuGlnGluGlnLeuHisLysAlaAsnGluLeu	406
QY	1087	CAGCTTTTAGTGGGAATAT-----GAAGTA	1113
DB	407	GluPheLeuHisAspGlnIleSerArgMetAsnGluGlyLysAsnPheGluAspIle	426
QY	1114	GCCATGCCAG-----GAAATGGAAGAATGTCCTCAATAAGCAAGAAAGACACACA	1161
DB	427	MetLeuGlnPheArgSerLeuGluGluArgAspValLeuSerLysLeuGlnThr	446
QY	1162	TGGGAGACTATTCTTGTGAGGAGAGGCTGCTCCATTCGAAACATTTTCTCAGGAGCT	1221
DB	447	LeuGluAspAspAsnAsnSerLeuArgLeuMetThrSerSerLeuGlyAsnGlnIleGlu	466
QY	1222	ACGTTGCGAGTCTCTCTTGGTGGACAGAG-----	1254
DB	467	SerLeuArgThrGlnAsnArgGluIleAspGluLysAsnHisLeuArgLeuAla	486
QY	1255	-----ACCAATGATAAATCTACGGCT-----CCTATTGCCAAA	1287
DB	487	SerLysAsnSerAspLysAlaLeuAlaGluThrAsnIleArgLeuGlnGluValThrLys	506
QY	1288	CCTCTGCCACT-----AGAAATTCAGAGACTCTCCAT-----	1320
DB	507	GluLeuGluThrLeuArgMetLysAsnSerAsnAspLeuAsnGluIleHisAspLeuArg	526
QY	1321	CAGGAAACAAAGCCCTGTTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCTTG	1380
DB	527	GluGluAsnGluGlyLeuThrLysLysIleAspSerIleThrLysGluLysAspArgLeu	546
QY	1381	ACTACAGATCTCAACAAAGA-----AAAGAAAGGAT	1413

547 IleAsnGluLeuGluGlnArgIleYssSerTyGluValAsnValSerGluLeuAsnGly 566  
1414 ACTCCAAATGAACCGCAAAATTAAGA-----ATATTT 1449  
567 ThrIleAspGluTyArgAsnLysLeuLysAspLysGluGluThrTyAsnGluValMet 586  
1450 TATCAGTTTCTCTAT---AACCAACAATCAAGCAACAACTGAAGCAAGAGATCAGCTG 1506  
587 AsnAlaPheGlnTyLysAspAsnAspLeuArgAgPheHisGluSerIleAsnLysLeu 606  
1507 CATTCGCCCTTGCTGACTCTGACGCGCAACTTATAGTTTACTCAAGCATCTTAAA 1566  
607 GlnAsp-----ArgGluLysGluLeuThrSerAsnLeuGluLys-LysAs 621  
1567 CTCTGCCATAGCAGATTTATCTTCAATATGTTTATCATCCAAAAGGTGCTAGGATAGAT 1626  
621 neuValIleSerSerLeuArgGluThrValAlaMetLeuGluLys----- 636  
1627 GTTCTTATCAATGAGTGTATGATGCTCTATGAGGAAATCTTCAGATATTCATCGC 1686  
637 -----GluArgGluSerIleLysLysTyLeuSe 646  
1687 C-----AACCTGGATTGCTTTAGTCGCAACGGACCCAG----- 1720  
646 rGlyAsnAlaLysAspLeuAspAsnThrAsnLeuMetGluIleLeuAsnAspLysIleSe 666  
1721 -----TTAAGAGAACCTTATCACACATA----- 1744  
666 rValLeuGlnArgGlnLeuThrAspValLysAspGluLeuAspValSerGluGluLys 686  
1745 -----TTCCTGTGTGAGGCGCAACGAAACGAAAGCAGCATGCTGAATTTCTGAA 1797  
686 gGluGluAlaIleValAlaGlyGln--LysLeuSerAlaSerPhe--GluLeuMetSer 704  
1798 TCTGAAGATGGGAGTAGACAGACAGAAACATATAGTAGGCGCAACATGCTGCTGAT 1857  
705 AsnGluLysGlnAlaLeuGluLys-----TyrSerSerLeuLysAsnGluLeu-- 721  
1858 TTCCATAGTAGTACTGCTTACTCTCCGTCACAGAAAGAAAGTAGTAGTGAAGAT 1917  
722 IleAsnAlaGlnAsnLeuLeuAspArgArgGluGluLeuSerGluLeuSerLysLys 741  
1918 GAAAGGATCCTGAATGGCTAGAGAAACCAATACACAAATGAAGAGTTTCTCAT 1977  
742 LeuPheGluGluArgLysIleArgSerGlySerAsnAspAspIleGluLysAsnLysGlu 761  
1978 GTTAATGAAGAGAGAGAAAGATGATG---AAACTCTGGAATCTC---CATGTCATGAAG 2031  
762 IleAsnValLeuAsnSerGluLeuAlaAspLysLeuAlaGlnIleArgHisLeuGluSer 781  
2032 CATGGTTTATGCTGACAAATCAATGATCATGCTGATGCTGTTGTAGAA---AAT 2088  
782 AspLysMetGluLeuAspLysLeuValHisHisLeuAsnArgGlyIleGluAlaAsn 801  
2089 TATGACAGAAAATTAATTAAGAGAAATTTATGTCGAAACTTCATGCTTCACTAGTCAGC 2148  
802 IleGluGluAsnAlaValLysLysArgLeuCys-----LeuLeuMet 815  
2149 ATGCATAGCTTAATCTTATAGCAATAGTCAATA----- 2184  
816 GlyCysAspTySerSerValSerIleLeuGlnIleValSerGlnIleGluHisPheVal 835  
2185 GATAAGCTGTTACCAAGCTCCGTAATGCGACGAAAATTAAGAA----- 2229  
836 AsnGlnGlnIleGlnThrIleArgSerLeuLysGlnGluLeuArgHisAspPheValGln 855  
2229 ----- 2229  
856 PheSerGlyLysLysGlnGlnGluLeuSerA:SerPheGluLysPheGlyLeuGlyThr 875  
2230 -----AAGGGGAATCTGCTTCCCTGCAACGAGAGAAATTAACGAGAACAAAT--- 2280  
876 GluThrLysHisAspLeuLeuAlaGlnArgAsnArgAsnValSerGluLysMetAsnAsp 895

2281 -----GGGACAGCAAAATGGATT--- 2298  
896 LeuGluAsnAlaAlaGlnLysPhePheSerSerProAspArgLysAsnGlyTyLeuTyr 915  
2299 -----ACTGAATTAACATAAAGAGAAAGCTTTGGAA----- 2331  
916 ProSerGluHisThrSerLysIleGluTyLeuGluLysThrIleGluAspLeuLysLeu 935  
2332 -----ACAGATAGTCTCTCAGGGTTTCA 2355  
936 AlaLeuGlnAspGluLeuLysAsnArgAsnLeuMetAspLysSerSerTyAsn 955  
2356 AAACAGAGCAAAAAA 2370  
956 LysGlnThrThrLys 960  
RESULT 14  
YCZ7\_YEAST  
ID YCZ7\_YEAST STANDARD; PRT; 1755 AA.  
AC P47058; P87194;  
01-FEB-1996 (Rel. 33, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transposon Ty1 protein B.  
GN TY1B OR YJ027W OR J1360.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
[1]  
SEQUENCE FROM N.A.  
RP MEDLINE=96109930; PubMed=8619316;  
RA Zagulski M., Babinaka B., Gromada R., Migdalski A., Rytko J.,  
RA Sulicka J., Herbert C.J.;  
RT "The sequence of 24.3 kb from chromosome X reveals five complete open  
RT reading frames", all of which correspond to new genes, and a tandem  
RT insertion of a Ty1 transposon."  
RL Yeast 11:1179-1186(1995).  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z49526; CAA89553.1; --  
DR MEROPS; A11.003; --  
DR SGD; S0003788; YJ027W.  
DR InterPro; IPR001969; Aspprotease AS.  
DR InterPro; IPR001042; Pept\_A113\_Ty1.  
DR InterPro; IPR001584; Rve.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF01021; TYA; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR Transposable element; Hypothetical protein; Hydrolase;  
KW Aspartyl protease; ATP-binding.  
FT ACT SITE 461 461  
FT NP\_BIND 1631 1638 ATP (POTENTIAL).  
SQ SEQUENCE 1755 AA; 198615 MW; 04E248A77FD7596F CRC64;  
Alignment Scores:  
Pred. No.: 0.00315 Length: 1755  
Score: 155.50 Matches: 159  
Percent Similarity: 36.70% Conservative: 99  
Best Local Similarity: 22.62% Mismatches: 232  
Query Match: 2.02% Indels: 213  
DB: 1 Gaps: 44  
US-09-874-162A-7 (1-4409) x YCZ7\_YEAST (1-1755)





```

QY 476 TCATAGCACAATATTTTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCCAGAA 535
Db 157 -----
QY 536 CAAACATCAAAAGGAAACATTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGA 595
Db 157 -----
QY 596 AAGGAGAGCAAGATCTCATAGCTTGTGAGCTCATTTGAGCTTACGTTTACTGTTTCT 655
Db 158 -----
QY 656 TCCACAAAATGATAAGCCATCACCAAACTCAGAAAATGAAACAAAATTCCTGTACCCCTGG 715
Db 161 roProGlyAlaProProProProProProSerSerGly-----ThrSerAlap 161
QY 716 AAGTCCTGCTTGTGAAGTTTGCCACAAAAGAAAGGATGTAAGTTGTCCAAATAAGGC 775
Db 174 -----ValProThrThrProProGlnAlaGlyGlyProProProProAlaA 190
QY 776 AAGTTCCC-----ACAGTAAAGACAGGTCCT----- 804
Db 190 laValProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyG 210
QY 805 --TTGATTCCTGACCTCAATCAAAACAAACCCGAAATTTCCCTCCCTTCAGTTTCCA 862
Db 210 lyLysMetPro-----GlyGlyProLysProGlyGlyGlyProGlyLysSerThrProG 228
QY 863 GTAATGAATTTGAACCTAGTAACCCATATGGTGAAGCTTACTCGTTGCTATTAGAG 922
Db 228 lyGlyHisProLysProHisArgGlyGlyGlyGlu----- 240
QY 923 TGACTCGTCCAGGAAGAGAGAGATT----- 948
Db 241 --ProArgGlyGlyArgGlnHisHisProProTyHisGlnGlnHisGlnGlyProp 260
QY 949 --AATGGAATGATTAATGGAAACCAATGAATAATATTGATGTCAATGAAGAGCTTCCAG 1006
Db 260 roProGlyGlyProGlyGlyArgSerGluGlyLysSerAspSerGluGlyPheLysA 280
QY 1007 CCAGAAAGAAACGAAATCGTGAGGATGGGAAAGACATTTGTCACAAATGACACTAT 1066
Db 280 laAsnLeuSerLeuLeuArgArgProGlyGlyLysThrTyThrGlnArgCysArgLeuP 300
QY 1067 TTGATAAAACAGGCGCTTACAGCTTTTAGATGGGGAATAT----- 1107
Db 300 heValGlyAsnLeuProAlaAspIleThrGluAspGluPheLysArgLeuPheAlaLysT 320
QY 1107 ----- 1107
Db 320 yrGlyGluProGlyGluValPheLeAsnLysGlyLysGlyPheGlyPheIleLysLeuG 340
QY 1108 -----GAAGTAGCCATGCGAGAAATGGAAGAAATGTCCAATAAGCAAGA 1150
Db 340 luSerArgAlaLeuAlaGluIleAlaLysAlaGluLeuAspAspThrProMetArgGlyA 360
QY 1151 AAAGA-----GCAACATGGGAGACTATTCTTGATGGGAAGAGGCTGCCTC 1195
Db 360 rgGlnLeuArgValArgPheAlaThrHisAlaAlaLeuSerValArgAsnLeuSerP 380
QY 1196 CATTC-----GAAACATTTTCTCAG-----GGACCTACGTTGCAGT 1231
Db 380 roTyValSerAsnGluLeuLeuGluGluAlaPheSerGlnPheGlyProIleGluArgA 400
QY 1232 TCACCTCT-----CGTGGACAGGAGAG-----ACCAATGATAAATCTA 1270
Db 400 laValValIleValAspArgGlyArgSerThrGlyLysGlyIleValIleGluPheAlaS 420
QY 1271 CGGCTCCTATTGCCAAACCTCTTCCCACTAGAAATTCAGAGAGT-----CTCCATC 1321
Db 420 erLysProAlaAlaArgLysAlaPheGluArgCysSerGluGlyValPheLeuLeuThrT 440

```

```

QY 1322 AGGAAAACAGCGCTTTCAGTTTAACTTACTCAAACTATTGCTGTTAAAGAAATCATTTGA 1381
Db 440 hrThrProArgProValIleValGluProLeuGluGlnLeuAspAspGluAspGlyLeuP 460
QY 1382 CTACAGATCTACAAAACAAGA-----AAAGAAAAGGATACTCCAAATGAAA 1426
Db 460 roGluLysLeuAlaGlnLysAsnProMetTyxGlnLysGluArgGluThrProProArgP 480
QY 1427 ACCGACAAAATTAAGAATATTTTATCAGTTTCTC-----T 1462
Db 480 heAlaGlnHisGlyThrPheGluTyxGluTyxSerGlnArgTrpLysSerLeuAspGluM 500
QY 1463 ATAACAACAATAACAGGCCAACAACTGAA-----GCAAGAGATGACCTG 1506
Db 500 etCluLysGlnGlnArgGluGlnValGluLysAsnMetLysAspAlaLysAspLysLeu 519

```

Search completed: August 25, 2004, 18:44:25  
Job time : 154 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 25, 2004, 18:24:39 ; Search time 445 Seconds

(without alignments)  
6252.224 Million cell updates/sec

Title: US-09-874-162A-7

Perfect score: 7690

Sequence: 1 cccgcgcgcgcgcgcag.....tacataaagtccaataatat 4409

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USPTO.spool\_p/US09874162/runat\_25082004\_171844\_2180/app\_query.fasta\_1.4551  
-DB=SPTRMBL\_25 -QWNT=fasan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09874162 @CGN\_1\_1\_717 @runat\_25082004\_171844\_2180 -NCPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DRT\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL\_25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvivirus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244.5	3.2	626	10 Q84UX5	Q84ux5 zea mays (m

ID	Q84UX5	PRELIMINARY;	PRT;	626 AA.
AC	Q84UX5;			
DT	01-JUN-2003 (TrEMBLrel. 24, Created)			
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	VEF family protein.			
GN	VEF101.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	PACCAD clade; Panicoideae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. B73;			
RA	Chandler V.L., Kaeppler S.M., Kaeppler H.F., Cone K.C.;			
RT	"Sequences from the Plant Chromatin Consortium (NSF Plant Genome program Grant 9575930).";			
RL	Submitted (AFR-2003) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. B73;			
RA	Bergstrom D.E., Springer N.M., Schmitt L., Guthrie E., Sidorenko L.,			

## ALIGNMENTS

### RESULT 1

Q84UX5  
ID Q84UX5  
AC Q84UX5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE VEF family protein.  
GN VEF101.

OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. B73;

RA "Sequences from the Plant Chromatin Consortium (NSF Plant Genome program Grant 9575930).";

RT Submitted (AFR-2003) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. B73;

RA Bergstrom D.E., Springer N.M., Schmitt L., Guthrie E., Sidorenko L.,

Q9V9Z6 drosophila  
Q7XSM1 oryza sativ  
Q8IBL3 plasmodium  
Q813N9 plasmodium  
Q81EH5 plasmodium  
Q96216 plasmodium  
Q81IC8 plasmodium  
Q59904 kluyveromyc  
Q8IB9 plasmodium  
Q8ICR3 plasmodium  
Q86H41 dictyosteli  
Q8IM18 plasmodium  
Q8IKP5 plasmodium  
Q81OT3 mus musculu  
Q81525 Plasmodium  
Q81K96 Plasmodium  
Q95151 canis fami  
Q8IC7 plasmodium  
Q9U4X0 plasmodium  
Q9AH10 borrelia bu  
Q81K4 plasmodium  
Q8AH7 borrelia bu  
Q9AH6 borrelia bu  
Q86V2 homo sapien  
Q95QW6 caenorhabdi  
Q8HY4 bos taurus  
Q81J8 plasmodium  
Q81L77 plasmodium  
Q81P4 plasmodium  
Q9AH8 borrelia bu  
Q51228 borrelia bu  
Q81B6 plasmodium  
Q8MK19 canis fami  
Q81V3 plasmodium  
Q81I50 plasmodium  
Q861V6 dictyosteli  
Q9W4G6 drosophila  
Q81J9 plasmodium  
Q9V7G8 drosophila  
Q8MLD9 drosophila  
Q12141 saccharomyc  
Q8IH9 plasmodium  
Q81J6 plasmodium  
Q77320 plasmodium

RA Sellinger D., Kaepler S.M., Cone K.C.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY232824; A084022.1; -;  
 DR InterPro; IPR007087; Znf C2H2.

DR PROSITE; PS00028; ZINC FINGER\_C2H2\_1; 1.  
 SQ SEQUENCE 626 AA; 71401 MW; 604C0D46BD5A8718 CRC64;

Alignment Scores:  
 Pred. No.: 1,458-10 Length: 626  
 Score: 244.50 Matches: 123  
 Percent Similarity: 34.21% Conservative: 110  
 Best Local Similarity: 18.06% Mismatches: 251  
 Query Match: 3.18% Indels: 197  
 DB: 10 Gaps: 22

US-09-874-162a-7 (1-4409) x Q84UX5 (1-626)

QY 436 GAGCCAAACACAGATCTATAGATTTCTCGAACTCGAATCTCATAGACCAATATTTTGTG 495  
 DB 53 LysProValGluLeuTyAsnIleIleGlnArgAlaMetLysAsnProLeuPheIle 72  
 QY 496 CACAGAACTCTTACTTACATCTCTCATCGAACTCCAGAACAAACATCAAGAGAAACA 555  
 DB 73 GlnArgCysLeuLeuTyAsnIleHisAlaArgLysLysArgIleGln----- 89  
 QY 556 TTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAAATGAAAGAGAGACGAATCTCAT 615  
 DB 89 ----- 89  
 QY 616 AGCTGTGAGCTCATTTGCGAGTTAGCTTACTGTTTCTTCTCCACAAAATGATAAGCCA 675  
 DB 90 -----IleThrIleSerLeuSerGly----- 96  
 QY 676 TCACCAAACTCAGAAAATGAACAAATCTGTT---ACCCTGGAAGCTCTGTTGTGAAA 732  
 DB 97 SerThrAsnThrGluLeuGlnThrHisTyValPheProLeuTyValLeuLeuAlaArg 116  
 QY 733 GTTCCGCCAATAAAGAAAGATGTAAGTTGTCAATAGCGCAAGTTCCACA----- 786  
 DB 117 ProThrSerAsnLeuSerLeuGluGlyHisSerProIleTyArgPheSerArgValCys 136  
 QY 787 -----GGTAAAGAGCAG-----GTG 801  
 DB 137 LeuLeuThrSerPheSerGluHisGlyAsnLysAspAsnSerGluAlaThrPheIleIle 156  
 QY 802 CTTTGTATTCCTGACCTCAATCAACAAACCCGGAATTTCCCGTCCCTTGCAGTTTCC 861  
 DB 157 ProAspValLysSerLeuSerThrSerArgAlaCysAsnHisAspIleIlePheIleSer 176  
 QY 862 AGTAATGAATTTGAACCTAGTAACAGCATATGTTGAAGTCTTACTCGTCTATTAGA 921  
 DB 177 CysGlyGlnVal----- 180  
 QY 922 GTGACTCGTCCAGGAAGAGAGAGTTTAATGGAATGATTAATGAGAAACCAATGAAAT 981  
 DB 181 -----GlyGlnSerAsnGlyGluAspAsnCysSer 190  
 QY 982 ATTGATGTCATGAAGAGCTTCCAGCCAGAGAAACGAATCTGAGGATGGGGAAG 1041  
 DB 191 -----GlyAsnHisValGluAspSerSer--- 198  
 QY 1042 ACATTGTTGCACAAATCACAGATTTTGTATAAAACAGGCGCTTACAGCTTTTGTAGATGGG 1101  
 DB 139 -----LeuGlnMetLeuGluGly 204  
 QY 1102 GAATATGAAGTAGCCATGCGAGAAATGAAGATGTCCATGAAGAAAGAGCAACA 1161  
 DB 205 -----LysCysSer 207  
 QY 1162 TGG-----GAGACTATTCTTGATGGGAAGAGGCTGCCCTCCATTGCAAAACATTTCT 1212  
 DB 208 TrpGlyLysIleProThrAsnLeuLeuAlaSerSerLeuGluSerCysValAsnLeuSer 227

QY 1213 CAGGGACCTACGTTGCGATTCACTCTTCGTTGGACGAGGACCAATGATAAATCTACG 1272  
 DB 228 LeuGlyHisIleValLeuLeuAlaSerLysValThrMetArgProSer----- 243  
 QY 1273 GTCCTTATTGCCAAACCTCTTCCCACTAGAAATTCAGAGAGTCTC-----CAT 1320  
 DB 244 -----PheLeuGluProLysPheLeuGluGlnAspSerCysLeuThrPheCysSerHis 261  
 QY 1321 CAGGAAAACACAGCTGGTTTCACTTAACTACTCAAACTATTGCTGTTAAAGAAATCATG 1380  
 DB 262 LysValAspAlaValGlySerTyLysLeuGlnLeuCysMetSerAlaGlnIleAlaGly 281  
 QY 1381 ACTACAGATCTA-----CAAACAAGAAAAGAAAGATATCTCCAAATGAA 1425  
 DB 282 AlaArgAspMetSerLeuSerProTySerSerTyAsnAspValProProSer 301  
 QY 1426 AACCGACAAAATTA-----AGAAATTTTATCATGTTTCTCTATTAAC 1467  
 DB 302 SerLeuSerAspIleIleArgLeuArgSerGlyAsnValLeuPheAsnTyLysTyLys 321  
 QY 1468 AACAAATACAGGCAACAAACTGAAGCAAGAGATGACCTGCATTGCCCTTGGTGTACTCTG 1527  
 DB 322 AsnAsnThrMetGlnGluThrGluValThrGluAspPheSerCysProPheCysTyLys 341  
 QY 1528 AACTGCCGCAAACTTTATAGTTTCTCAAGCATCTTAACTCTGCCATAGAGATTTATC 1587  
 DB 342 ArgCysGlySerPheLysGlyLeuGlyCysHisLeuAsnSerSerHisAspLeuPheHis 361  
 QY 1588 TTCACATGTTTATCATCCAAAAGGTCTAGATAGATGTTTCTTATCAAT---GAGTGT 1644  
 DB 362 TyrGluPheTrpIleSerGluGluTyGlnValValAsnValSerLeuLysAlaAspAla 381  
 QY 1645 TATGATGGTCTCTATCGAGAAATCTCAGGATATTTCATCGCAACCTGGATTTCTCTTT 1704  
 DB 382 TrpArgThrGluLeuPheAlaGluGlyValAspProArgHisGlnThr---PheSerTyL 400  
 QY 1705 -----AGTCCCAAGGACCCAGTTAAG---AGAACACCT 1734  
 DB 401 ArgSerArgPheLysLysArgArgSerLysAsnThrMetGluLysIleArgHisVal 420  
 QY 1735 ATCACACATATTCTTGTGTGC---AGGCCAAACCAACAAAGCAAGCATGTCTCAATTT 1791  
 DB 421 HisSerHisIleMetGluSerGlySerProGluAspGluAlaGlySerGluAspAsnPhe 440  
 QY 1792 CTTGAATCTGAAGATGGG----- 1809  
 DB 441 ValGlnGlyGluAsnGlyThrSerValAlaAsnAlaSerIleAspProAlaGlnSerLeu 460  
 QY 1809 ----- 1809  
 DB 461 HisGlySerAsnLeuSerProProThrValLeuGlnPheGlyLysThrArgLysLeuSer 480  
 QY 1810 GAAGTAGAAGACAAAGAACATATAGTAGTGCCCAATCGTCTGTATTTCATAGTAT 1869  
 DB 481 GluArgSerAspProArgAsnArgGlnLeuLeuGlnLysArgGlnPhePheHisSerHis 500  
 QY 1870 ACCTGCTTACCTCTCCGTCACCAAGAAATG-----GAAGTAGATAGTCAAGATCAAAAG 1923  
 DB 501 ArgAlaGlnProMetGlnLeuGlnValPheSerAspArgAspSerGluAspGluVal 520  
 QY 1924 GATCTCTGAATGGCTAAGAGAAAACCAATACACAAATTTGAAGATTTTCTGATTTAAT 1983  
 DB 521 AspAspAspIleAlaAspPheGluAspArgArgMetLeuAspAspPheValAspValThr 540  
 QY 1984 GAAGAGACAAAGAGTGTGTAATCTCGAATCTCCATGTCTCAATGAACATGGGTTTATT 2043  
 DB 541 LysAspGluLysLeuIleMetHisMetTrpAsnSerPheValArgLysGlnArgValLeu 560  
 QY 2044 GTGCAATCAATCAATCATGCTGTATGCTGTGTTGTAGAAAATTTAGACAGAAAAATA 2103  
 DB 561 AlaAspGlyHisIleProTrpAlaCysGluAlaPheSerGlnLeuHisGlyArgGlnLeu 580  
 QY 2104 ATTAGAAG---AATTATGTCGAAACTTCATCTCTAGTCAGCATGCATCACTTT 2160



Db 581 IleglnAsnProAlaLeuLeuTrpGlyTrpArgPheMetIleIleIleLeuTrpAsnHis 600  
QY 2161 AATCTTATTAGCATATGTCATAGATAAAGTGTACCAAGCTCGGCAAGTGCAGAA 2220  
Db 601 ASnileuAspAlaArgThrMetAsnThrCysAsnThrValLeuGlnIleLeuGlnGlu 620  
QY 2221 AAA 2223  
Db 621 Glu 621  
RESULT 2  
Q9V9Z6 PRELIMINARY; PRT; 628 AA.  
ID Q9V9Z6 Q960L0;  
AC Q9V9Z6; Q960L0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG12054-PA (L046863p).  
GN CG12054;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flooker A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Submitted 287:2185-2195 (2000).  
[2]  
SEQUENCE FROM N.A.  
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacle B.J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradycky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A.  
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
[6]  
SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarn H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacle B.J., Paragas V., Park S., Phouanenavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celnik S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003777; AAF57132.2; -;  
DR EMBL; AY052003; AAK93427.1; -;  
DR FlyBase; FBgn0039831; CG12054.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00036; znf\_C2H2\_3.  
DR SMART; SM00355; Znf\_C2H2\_3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 1.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 628 AA; 66610 MW; 860C67EA04D41CF5 CRC64;  
Alignment Scores:  
Pred. No.: 3,51e-09 Length: 628  
Score: 227.00 Matches: 143  
Percent Similarity: 34.57% Conservative: 81  
Best Local Similarity: 22.07% Mismatches: 208  
Query Match: 2.95% Indels: 216  
DB: 5 Gaps: 26  
US-09-874-162a-7 (1-4409) x Q9V9Z6 (1-628)  
QY 70 GCCTCTCTCTCTCTCAATGCTCCGAGTTCGGGGCTGGGACTCCACTTCCACCCCTG 129  
Db 2 AlavalPheLeuLeuLeuValCysLysTyrAsnGlyCysGlyLeuThrPheProSerLeu 21  
QY 130 GCGAGCTCTATCGAGCATCTCGAGGACACACCACATCGATACAGATCCACGGGTTTAGAA 189  
Db 22 SerAspLeuIleSerHisIleGluAspThrHisIleAspTyrAspProLysValValGlu 41  
QY 190 AAAAAGAATTACAGCAGCAACCTATGTTGCCCTGAGTTACATAATAGATTTCATGACA 249  
Db 42 GlnIleGluGlnAlaGlnProAlaCysLeuProLeuSerTyrValLeuArgPheIleThr 61  
QY 250 GATGTCGCGCGAGGAGGAGGAGTCCCTAAAGAGAGAGATTACGCCGAGCTCTCGCTG 309



Alignment Scores:			
Pred. No.:	1,75e-07	Length:	614
Score:	205.50	Matches:	135
Percent Similarity:	34.55%	Conservative:	111
Best Local Similarity:	18.96%	Mismatches:	263
Query Match:	2.67%	Indels:	203
DB:	10	Gaps:	31
US-09-874-162A-7 (1-4409) x Q7XSM1 (1-614)			
QY	409	TCCTTATCCGAGCAGCACTCCGACAG	-----CCAACACAG 447
DB	11	SerSerAspGlnGlnLeuAsnProGluGluAsnLeuAlaLeuTyrCysLysProLeuGlu 30	: : : : :     : : :
QY	448	ATCTATAGATTTCTTCGAACCTCGGATCTCATAGCACCACCAATATTTTGCACAGCACTCTT	: : : : :     : : :
DB	31	LeuTyrAsnPhelIeAlaGlnGluAsnProTyrLeuGlnArgCysLeu 50	: : : : :     : : :
QY	508	ACTTACATCTCTCATCGAACTCCAGAACAAATCAAAAGAAACATTTAAAGTTGAT 567	: : : : :     : : :
DB	51	LeuTyr	: : : : :     : : :
QY	568	GATATGTTATCAAAAGTAGAGAAATGAAGAGAGAGCAAGATCTCATAGCTTGTGACGT 627	: : : : :     : : :
DB	53	-----LysIleArgAlaLysGlnLysLys-	: : : : :     : : :
QY	628	CATTTCAGCTTACGTTTACTGGTTCTTCCACAAATAATGAPACCCATCACCAACTCA 687	: : : : :     : : :
DB	61	ArgIleGlnIleThrIleSer	: : : : :     : : :
QY	688	GAATATGAA-----CAAAATCTGTACCTCGAAGTCTCTCTGTGNAAGTT 735	: : : : :     : : :
DB	72	AsnAsnLysGluLeuGlnAlaGlnAsnIlePheProLeuTyrValPheAlaArgPro 91	: : : : :     : : :
QY	736	TOCCACAAAAAAGAAAGATGAAGTTGTCCAAATAGGCAAGTTCCACACAGGTAATAAG 795	: : : : :     : : :
DB	92	ThrSerAsnValProIleGluGlyHisSerProIleTyrArgPhe	: : : : :     : : :
QY	796	CAGTGCCTTTCATCTCCTGACCTCAATCAACAAACCCGGAAT-----840	: : : : :     : : :
DB	108	GlnAlaArgLeuLeuThrSerPheAsnAspSer-----GlyAsnAsnAspArgAlaGlu 125	: : : : :     : : :
QY	841	-----TTCCCGTCCCTTCAGTTTCCAGTTTCCAGTAATGAATTGAACCTAGTAACAGC 888	: : : : :     : : :
DB	126	AlaThrPheValIleProAspLeu-----GluThrLeu 136	: : : : :     : : :
QY	889	CATATGGTGAAGTCTTACTCGTTGCTATTAGAGTACTCGTCCAGGAAGAGAGTTT 948	: : : : :     : : :
DB	137	IleAlaThrGlnAlaTyrGlyLeuThrPheIleLeuValSerArgGlyThrLysLys---155	: : : : :     : : :
QY	949	AATGGAATGATTAATGGAGAACCAATGAAATATTGATGTCATGAA-----GAGCTT 1002	: : : : :     : : :
DB	156	-----AsnLysGlyArgThrGlyGlnAsnLeuCysGluAsnAspCysSerGluLys 172	: : : : :     : : :
QY	1003	CCAGCCAGAGAAACGAAATCGTGAGGATGGGAAAGACATTTGTTGCACAAATGACA 1062	: : : : :     : : :
DB	173	HisValAspTyrSerSerLeuArgLysLeuAlaGlyLysCysPheTrpGlyLysIlePro 192	: : : : :     : : :
QY	1063	GTATTTGATAAACACAGCGCTTACAGCTTTAGATGGGAATTAAGTAGGCCATCGAG 1122	: : : : :     : : :
DB	193	IleThrLeuLeuAsnSerSerLeuGluThrCys-----AlaAspLeuIleLeuGly 209	: : : : :     : : :
QY	1123	GAAATGGGAAGATGTCCAAATGAAGAAAGAGACGACATGGGAGACTATTCTTGTGATGGG 1182	: : : : :     : : :
DB	210	HisIleValGluSerProIleSer-----IleCysMetSerProGly 223	: : : : :     : : :
QY	1193	AAGAGGTCCTCCATTCGAACATTTTCTCAGGAGCCTAGTTGCAATTCACCTCTTCGT 1242	: : : : :     : : :
DB	224	TyrLeuGluProThrPhe-----LeuGluHisAspAsnCysLeuSerPheCysSerArg 241	: : : : :     : : :
QY	1243	TGGACAGGAGAGACCAAGTAAATCTACGCTCTTATGTCACCAACCTCTTGCCACCTAGA 1302	: : : : :     : : :
DB	242	-----LysAlaAspAlaMetValProTyrGlnLeuGlnValLys 254	: : : : :     : : :
QY	1303	AATTCA-----GAGAGTCTCCATCAGGAAAAACAAG 1332	: : : : :     : : :
DB	255	ValSerAlaAlaGluAlaGlyAlaLysAspIleLeuLysSerProTyrAsnSerPheSer 274	: : : : :     : : :
QY	1333	CCTGTTTCACTAAACCTTAAAGTCTTAAAGTCTTAAAGTCTTAAAGTCTTAAAGTCTTAA 1392	: : : : :     : : :
DB	275	TyrSerAspValProProSerLeuLeuLeuArgIleValArgAlaLeuGlnTyrTyrAla 294	: : : : :     : : :
QY	1393	CAACACAGAAAGAAAGGATATCTCCAAATGAACCGCAAAATTAAGAAATATTTAT 1452	: : : : :     : : :
DB	295	AsnThrMetLys-----PhePhe 300	: : : : :     : : :
QY	1453	CAGTTT-----CTCTATAACAACAATACAAGCGCAACAACCTGAA 1491	: : : : :     : : :
DB	301	LeuTyrArgLeuArgValGlyAsnValLeuPheAsnTyrLysAsnThrGlnMetSerGlu 320	: : : : :     : : :
QY	1492	GCAAGAGATGACCTGCAATGCCCTTGTGTACTCTGAACCTGCCGCAAACTTTTAGTTTA 1551	: : : : :     : : :
DB	321	ValThrGluAspPheThrCysProPheCysLeuValArgCys-----334	: : : : :     : : :
QY	1552	CTCAGCATCTTAAACTCTGCCATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAA 1611	: : : : :     : : :
DB	335	---GlyAsnPheLysIleSerGluAspTyrGlnAlaValAsnValThrLeuLysLysAsp 353	: : : : :     : : :
QY	1612	GGTCTAGGATAGATGTTTCTTCAATGAGTGTATGATCGCTCTCATGAGGAAATCCT 1671	: : : : :     : : :
DB	354	AsnMetArgThrGluPheValAlaAlaGlu---ValAspAsnSerHis-----368	: : : : :     : : :
QY	1672	CAGATATTTCATGCCCACTGATTTGCTTTTAGTCGCAACCGA-----CCAGTTAAG 1725	: : : : :     : : :
DB	369	ArgIlePheTyrTyrArgSerArgPheLysLysSerArgThrGluIleLeuProValAla 388	: : : : :     : : :
QY	1726	AGAACACCTATCACATATTCTTGTGTGC---AGGCCAAAAACGAAACGAAAGCATG 1782	: : : : :     : : :
DB	389	ArgAlaAsp---AlaHisIleMetGluSerGlySerProGluLeuThrGlnAlaGluSer 407	: : : : :     : : :
QY	1783	TCTGAATTTCTGAATCTGAAGATGGGAGTA-----GAACAGCAAAACACATAT 1833	: : : : :     : : :
DB	408	GluAspAspValGlnGluAsnGluAsnAlaLeuIleAspAspSerLysLysLeuHis 427	: : : : :     : : :
QY	1834	AGTAGTGGCCAC-----1845	: : : : :     : : :
DB	428	GlySerAsnHisSerGlnSerGluPheLeuAlaPheGlyLysSerArgLysLeuSerAla 447	: : : : :     : : :
QY	1846	-----AATCGCTCTG-----TATTTCCATAGTAT 1869	: : : : :     : : :
DB	448	AsnArgAlaAspProArgAsnArgLeuLeuGlnLysArgGlnPheIleHisSerHis 467	: : : : :     : : :
QY	1870	ACCTGC-----1875	: : : : :     : : :
DB	468	LysAlaGlnThrProLeuGlyAspLysValThrAsnLeuValThrAlaAspAlaMet 487	: : : : :     : : :
QY	1876	---TTACCTCTCCGTCACAGAAATG-----GAAGTAGATAGTAGAAGTAAAGGAT 1926	: : : : :     : : :
DB	488	SerValProMetThrPheGluValLeuSerAspAsnAspSerGluAspGluValAsp 507	: : : : :     : : :
QY	1927	CCTGAATGGCTAAGAGAAAAACCATTCACAAATTAAGAGTTTCTGATGTTAATGAA 1986	: : : : :     : : :
DB	508	AspAspIleAlaAspLeuGluAspArgArgMetLeuAspAspPheValAspValThrLys 527	: : : : :     : : :
QY	1987	GGAGAGAAAGAGTATGAACTCTCGAATCTCCATGTCATGAGCATGGTTTATGCT 2046	: : : : :     : : :
DB	528	AspGluLysArgIleMetHisMetTrpAsnSerPheIleArgLysGlnSerIleLeuAla 547	: : : : :     : : :
QY	2047	GACATCAATGAATCATGCTGTATGCTGTTGTAGAAAAATTTATGCAGAAAAATAAT 2106	: : : : :     : : :
DB	548	AspSerHisValProTrpAlaCysGluAlaPheSerArgHisGlyGluGluLeuLeu 567	: : : : :     : : :
QY	2107	AAGAAAGAT---TTATGTGGAACCTTCATGCTTCATGTCAGTCAGTCAGTCATTAAT 2163	: : : : :     : : :
DB	568	GluAsnSerAlaLeuLeuTrpGlyTrpArgMetPheMetIleLysLeuTrpAsnHisSer 587	: : : : :     : : :

QY 2164 CTTATTAGCATATGCTCAATAGATAAAGCTGTTACCAAGCTCCGTGAATGCGACAAAA 2223  
 Db 598 LeuLeuSerAlaArgThrMetAspThrCysAsnArgIleLeuAspAspIle----- 604  
 QY 2224 TTAGAAAGGGGAATCTGCTTCCTCCCTGCAACGAA 2259  
 Db 605 -----LysAsnGluArgSerAspProLysLysGln 614  
 RESULT 4  
 Q81BL3 PRELIMINARY; PRT; 1605 AA.  
 AC Q81BL3 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN MAL7P1.127.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,  
 RA Quail M., Barrell B;  
 RL Submitted (SEP-2002) to the ENBL/GenBank/DBJ databases.  
 DR EMBL; AL844506; CAD50987.1; ..  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004672; F:protein kinase activity; IEA.  
 DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR000719; Prot\_Kinase.  
 DR Pfam; PF00069; pkinase\_1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1605 AA; 194644 MW; B63DF776A3E206A2 CRC64;

## Alignment Scores:

Pred. No.: 2,38e-06 Length: 1605  
 Score: 192.00 Matches: 263  
 Percent Similarity: 32.18% Conservative: 174  
 Best Local Similarity: 19.37% Mismatches: 429  
 Query Match: 2.50% Indels: 492  
 DB: 5 Gaps: 59  
 US-09-874-162A-7 (1-4409) x Q81BL3 (1-1605)  
 QY 508 ACTTACATGCTCATCGAACTCCAGAACACATCAAAAGGAAACATTTAAAGTTGAT 567  
 Db 233 ThrTyrLysGluAsnLysAsnValHis---TyrIleLysAsnLysLeuTyr-----Asp 249  
 QY 568 GATATGTTATCAAAAGTACAGAAATGAAGAGAGACGAAGAATCTCATAGCTTCAGCT 627  
 Db 250 HisValIleSerLysAsnGluMetTyrThrGlyGlnThrThrLysAsnAsnTyrLeuAsp 269  
 QY 628 CATTGCGAGCTACCTTACTGCTTCTTCAC----- 660  
 Db 270 LysLeuPheLysAsnPheIleTyrPheTyrHisIleGlnThrLeuThrAsnValSerTyr 289  
 QY 661 -----AAAAATGATAGCCATCACCACAACTCAGAAAT--- 693  
 Db 290 AspGluArgThrLysGlnTyrLysGlnAsnLysLysCysIleLysAspLysLeuAsnAsn 309  
 QY 694 -----GAACAAATCTGTTACCTGGAAGTCTCTGTTG 729  
 Db 310 IleThrAsnAsnTyrLysGlnTyrAspLeuAsnSerPheSerAspSerAlaLeuTyrIle 329  
 QY 730 AAA-----GTTTGGCCACAAA---AAAAGAAAGGATGAAGTTGCCA----- 768  
 Db 330 LysGlnLysIleIleCysSerLysGluLysArgLysThrIleAsnIleLysGlySerHis 349  
 QY 769 -----ATAAGGCAAGTTCCACAGGTAAA 792  
 Db 350 LysTyrLysAspLysLysAsnGlnGlnIleTyrAsnGluLysAsnAsnGluThrArgGln 369

QY 793 AAGCAGGTGCTTTGATTCCTGAGCCTCAATCAACAAACCCGGAATTTCCCGTCCCTT 852  
 Db 370 LysGlnIleProLeuGluSerTyrLysLysLysThrLysLysIleAsnLysGluGlnGln 389  
 QY 853 GCAGTTTCC-----AGT 864  
 Db 390 ThrIleGluGlnIleTyrGlnHisAlaLeuLeuValLysArgLysIleValAspIleGln 409  
 QY 865 AATGAATTGAACTAGTAACAGCATTATGGTGAAGTCTTACTGCTCTCTATTAGAGTG 924  
 Db 410 GlnGluGlnLysThrAspAsnLysTyrIleGluSerAsnTyrThr----- 424  
 QY 925 ACTCGTCCAGGAAGAAGAGAGATTAATGAATGATTAAATGCGAAAC----- 972  
 Db 425 -----AsnAsnValLysAsnAsnGlnThrAsnLysHisAsn 436  
 QY 973 -----AATGAAATATATGATGTCATGAAGAGCTTCCAGCCAGAAAGA 1014  
 Db 437 AlaMetAspProSerAsnAsnGluAsnIleSerSerThrSerLeuAsnValSerGluLys 456  
 QY 1015 AAACGAAATCGTACGATGGGAAAGACATTTGTCACAAATGACAGTATTT----- 1068  
 Db 457 LysMetPheArgPhePhePheLysThrSerGluLysAsnAspThrLeuPheLysIle 476  
 QY 1068 ----- 1068  
 Db 477 GlnAsnValPheIleLeuIleIleTyrIleLysLeuPheTyrIleTyrTyrLeuLys 496  
 QY 1069 -----GATAAAACAGCGCTTACAGCTTTTAGTGGGATATAG----- 1108  
 Db 497 TyrTyrLysAsnLysLysAsnLysGluIleSerLeuLeuAsn-ValAsnIleGlnAspIle 516  
 QY 1109 ----AAGTAGCCATGCAGG---AAATGGAAGATCTCCAATGAAGCAAGAAAGACACA 1161  
 Db 516 ePheLysArgLeuCysLysHisLysLeuAsnAsnTyrSerTyrProSerLysTyrArgHis 536  
 QY 1162 TGGAGACTATTCTTGTAGGGAGAGGCTGCCTCCATTCG----- 1201  
 Db 536 shisIleIleAsnLysLeuLysArgLysArgTyrAsnAsnIleIleTyrGluIleIleLys 556  
 QY 1202 -----AAACATTTCTCAGG-----GACCTACGTTGACGTTCACTCTCGT 1242  
 Db 556 sAsnIleIleGluTyrPheLeuAsnCysLysLeuAsnLeuSerPheIleLysLeuIleGln 576  
 QY 1243 TGGACAGGAGAGACCAATGATAAATCTACGGCTCTCTATTGCCAAACCTTTGCCACTA-- 1300  
 Db 576 uAsnMetTyrSerGluPhePheAsnIleHisLeuLeuLysTyrAsnLeuArgAsnIlePhe 596  
 QY 1301 ----GAAATTCAGAGAGTCTCCATCAGGAAACAGCCCTGGTTCAAGTAACTACTC-- 1354  
 Db 596 eTyrSerAspGlnAsnAspGlnAsnGlnLysThrLeuPheLeuAsnLeuLeuHis 616  
 QY 1355 -AAACTATTGCTGTTAAAGAATCATTTGACTACAGATCTACAAACAGAAAGAAAGGAT 1413  
 Db 616 sLysCysLeuLeuLeuAsnAsn----- 623  
 QY 1414 ACTGAAATGAAACCGACAAATAATTAAGATATTTTATCAGTTTCTCTATAACACAAAT 1473  
 Db 624 -LeuGlnCys--AsnSerThrSerLeuLeuSerHisTyrTyrPhePheTyrAsnValAsn 642  
 QY 1474 ACAAGCAACAAACTGAACAGAGATGACCTGCATTGCTGCTGCTGCTGCTGCTGCTGCTG 1533  
 Db 643 IleTyrLysHisAsnAsnCysLeuGluThrAsnTyrThrGluLysAspThrPheAsnIle 662  
 QY 1534 CGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCGCATAGCAGATTTATCTTCAAC 1593  
 Db 663 ProArgAsnMetSerLeuLeuGluAsnThrThrIleThrAsnGluLysValLeuMetAsn 682  
 QY 1594 TATGTTTATCATCCAAAGGCTGCTAGGATAGATGTTCTTATCAATCAGTGTGTTATGATGC 1653  
 Db 683 LysValSerHisIleAsnAsnPheIleSerAsnAlaCysThrSerHisIleTyrThrThr 702



```

Db 1359 IleSerIleLeuMetTyrLeuLysTyrLeuIleAsnIleThrThrAsnGluAsnIle 1378
QY 3193 -----ATTAAGCTGTTCATATCTTTAAATAGGATTGCAACCAAGG 3234
Db 1379 TyrLysLysValPheSerLeuSerAsnLeuIleAsnLeuAsn-LeuIleLeuAsnThrSe 1398
QY 3235 AAAGAACCCATTGAGATTTAAGATG-----CACTT 3267
Db 1398 rArgTyrLeuPheGlu-----ArgCysProLysTyrTyrLeuLeuPheProPheHisLe 1416
QY 3268 ATAAGGGAGAGAGTCTTCTTAAAGAGTCAACAGCAAAACTGTTATGCTTTTATTGTTT 3327
Db 1416 uGlnIleGluGlySerSer-----IleAsnGluLysSerIle----- 1428
QY 3328 GCAAGGATGCTTTGTAAGTGTTCATGAATAGATATCAATCAATAGAGATAAGCTGACTT 3387
Db 1429 -----LeuPheAspLysGlyGlyAsnGluAsnIleSer----- 1439
QY 3388 GAATCATTTTGAGCAATTTGCCCTGTTATGTTATGTTTCAACCCACATATTTCAGCTTG 3447
Db 1440 -AsnHis-----IleAsnTyrIleSerHis----- 1447
QY 3448 GATTTCCTCCACAGAAAGTGGATTCACTACTGG 3481
Db 1448 -LeuIleSerAsnLysLysTrpIleArgTyrTyr 1458

RESULT 5
Q813N9 PRELIMINARY; PRT; 9341 AA.
AC Q813N9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF01120W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1] _SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagals K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lemard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929353; CAD51588.1; -
KW Hypothetical protein.
SQ SEQUENCE 9341 AA; 1121395 MW; BAE88CDEDF66B92D CRC64;

```

```

Alignment Scores:
Pred. No.: 6.52e-06 Length: 9341
Score: 188.00 Matches: 196
Percent Similarity: 33.16% Conservative: 179
Best Local Similarity: 17.33% Mismatches: 391
Query Match: 2.44% Indels: 365

```

```

DB: 5 Gaps: 48
US-09-874-162A-7 (1-4409) x Q813N9 (1-9341)
QY 106 TCGGAGCTACCTTCCCGACCTTCCCGACCTCATCGAGCAGCATCGAGGACAAACCAATC 165
Db 2810 CysAsnArgHis-----LeuSerAsnTyrMetAsnSerIleLeuAsnAsnSerAla 2826
QY 166 GATACAGATCCACGGGTTTAAAGAAAACAAGAATTACAGACCCCAACTATGTTGCCCTG 225
Db 2827 AspAsnLysHisAsnAsnLeuAspLysGluArgGluLysGluProIle----- 2842
QY 226 ACTTACATAAATAGATTATGACAGATGCTGCCCGCAGAGAGAGGAGTCCCTAAAGAG 285
Db 2843 ---PheMetAsnProTyrGluAsp-----LysAsp 2851
QY 286 AAGATTCCCGAAGCTCTCGCTGACTCTGTCTCAGCTCAGTGTCTCGA-----GGGAAT 339
Db 2852 LysThrLysAspValGluHisSerIleAsnSerAsnGlnIleArgTyrValGlyAsn 2871
QY 340 GTGTCCACTCCCGCAGGCACAGCAGTGGAGCTTACTCCCGCTGACCCCAATC 399
Db 2872 ValLysCysIleGlyAsnAsnLysHisLysTyrValIleProPheLysSerAsnVal 2891
QY 400 ACCCGCTCTCTTCATTCGCGACGAGCAGCTCCGACAGAGCAACACAGATCTATAGATT 459
Db 2892 -----LysAsnLysMetGluGlyIleSerSerSerSerSerCysIleTyrLysLys 2909
QY 460 CTTGCACTCGGAATCTCATGACCAATATTTTGCAC----- 498
Db 2910 ThrLysArgLysLysTyrIleAspAspIleAspCysAsnAsnIleAsnAspLysLys 2929
QY 499 AGAATCTTACTTACATGCTCATCGAAATCCACA-----ACAAATCAAAAGG 549
Db 2930 LysSerTyrAspIleLeuTyrGlyAsnAsnSerGlnGluTyrTyrGluAsnIleLysAsn 2949
QY 550 -----AAACATTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAAGGAGCAA 606
Db 2950 LeuArgAsnTyrLeuLysAsnAspMetIleLysAsnValLysLys-----SerSer 2966
QY 607 GAATCTCATAGCTTGTCTACCTCATTTGCAGCTTACGTTT----- 645
Db 2967 GluHisLysSerIleAspGlyAsnAlaAsnAsnLeuTyrAspIleProProAsnAspAsn 2986
QY 646 -----ACTGGTTTCTTCCACAAAATGATAGCCATCACCACAACTCAGAAAATGAA 696
Db 2987 GluMetLysAspGluPheLeuIleLysAsnAsp-----AspAspAsnLys 3001
QY 697 CAAATTCGTGTTACCTCGGAGTCTGTTGTGAAGTTTCCACAAAAGAAAGAGAT 756
Db 3002 MetAsnSerMetCysValAsp-----LysAspAspTyrLysGlnArgSerTyr 3017
QY 757 GTAAGTTGTCCAAATAAGG---CAAGTCTCCACAGGTAAAGAGCAG----- 798
Db 3018 IleSerTyrLeuLeuAsnTyrLysLysAsnGlyLysAsnGlyLysSerThrLysAsn 3037
QY 799 -----GTGCTTTGATTCTCGACCTCAATCAACAAACAAA 831
Db 3038 AsnIleMetLysArgPheAsnSerLeuIleHisPheIleHisAsnAsnThrSer 3057
QY 832 CCGGA-----AATTTCGC 846
Db 3058 LysGlyAlaSerLysLysIleLysGlyLysArgSerLeuSerLeuPheSerPheLys 3077
QY 847 TCCCTTGAGTTTCCAGTAATGAATTAAGTAACTAGTAAACAGCATATCGTGAAGTCTTAC 906
Db 3078 SerLeuProIleArgArgLysPheGlyLysTrpAsnSerAsnThrIleAspValPhe 3097
QY 907 TCGTTG-----CTATTTAGAGTGACTCGTCCAGGAAGAGAGAGTTAATCGA 954
Db 3098 GlnValGlyGlnGluAsnAsnTyrLysLysAsnLysLysLysThrLysLysPhePheGly 3117
QY 955 ATGATTAATGAGAAACC-----AATGAAATATTATGATGTCAT 993

```

3337	Asn	Phe	Met	Lys	Leu	Asn	Lys	Arg	Lys	Gly	Ser	Leu	Val	Cys	Cys	Tyr	Val	Pro	Val	Phe	3350
2008	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2022
3357	Tyr	Lys	Asn	Leu	Ser	Leu	Thr	Tyr	Arg	Asn	Leu	Tyr	Lys	Cys	Leu	Leu	Leu	Leu	Asn	Lys	3376
2023	GT	CAT	GA	G	CAT	GGG	T	T	T	T	T	G	C	T	G	C	A	A	T	G	2079
3377	Met	Glu	Ser	Leu	Pro	Tyr	Leu	Ser	Lys	Leu	Tyr	Lys	Asn	Glu	His	Met	Cys	Lys	Tyr	Leu	3396
2080	GT	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	2139
3397	Phe	Gly	Asn	Lys	Asn	Lys	His	Asn	Leu	Asn	Asp	Asn	Tyr	Leu	Lys	Glu	Tyr	Leu	His	3416	
2140	CT	AG	T	C	AG	C	AT	G	C	T	T	A	T	T	C	T	T	A	T	T	2199
3417	Leu	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	3423	
2200	A	AG	C	T	C	G	T	C	A	A	A	T	T	T	T	T	T	T	T	T	2247
3424	Glu	Lys	Glu	Gln	Tyr	Glu	Gln	Lys	Tyr	Ser	Phe	Glu	Leu	Lys	Lys	Asp	Leu	Phe	Lys	Asp	3443
2248	C	T	G	C	A	A	C	G	A	A	A	T	A	C	T	G	A	A	A	A	2307
3444	Met	Lys	Asn	Asp	Arg	Thr	Gln	Asn	Lys	Leu	Asn	Lys	Asn	Leu	Asn	Leu	Asn	Leu	Asn	3459	
2308	A	A	C	T	C	A	A	A	A	G	A	A	G	C	T	T	T	T	T	T	2355
3459	e	Lys	Glu	Lys	Arg	Asp	Tyr	Tyr	Asn	Asn	Ser	Asn	Leu	Asn	Leu	Leu	Asn	Met	Lys	Tyr	3479
2356	A	A	A	C	A	G	A	G	A	A	A	A	A	C	T	G	A	A	A	A	2415
3479	s	Lys	Asn	Asp	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	3499	
2416	A	A	T	T	C	A	T	T	T	T	T	T	T	T	T	T	T	T	T	2458	
3499	n	Lys	Asn	Asn	Tyr	Asn	Asn	Asn	Asn	Lys	Lys	Asn	Tyr	Ser	Tyr	Ser	Asn	Leu	Pro	3519	
2459	-	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2517	
3519	u	Leu	Ser	Asn	Asn	Leu	Lys	Gly	Lys	Gly	Asn	Lys	Tyr	Tyr	Gln	Leu	Phe	Gln	Ser	3539	
2518	A	T	A	T	T	T	C	A	T	C	A	G	G	T	T	C	T	A	---	---	2537
3539	l	ni	le	Val	Gln	Asn	le	Phe	Asn	Tyr	Arg	Asn	Lys	Asn	Lys	His	Asn	Asn	Val	3559	
2538	-	C	T	T	C	A	T	T	A	T	C	A	G	A	T	T	A	T	A	---	2595
3559	s	ni	le	Glu	Asn	Val	Met	Gln	Asn	Glu	Val	le	Lys	Asp	Asp	Val	Lys	Pro	Leu	3579	
2596	T	C	T	G	T	A	C	T	T	T	A	A	G	A	A	G	A	A	T	---	2633
3579	e	r	Cys	Ser	Leu	Tyr	Asp	Asn	Ser	Asn	Leu	Ser	Tyr	Pro	Phe	His	ile	Tyr	Asn	3599	
2634	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2679	
3599	s	n	Asn	Asn	Asn	Asn	Gln	Lys	Cys	Asn	Ser	Ser	Asn	Ser	Asn	Asn	Cys	Asn	Ser	3619	
2680	A	T	T	T	T	C	A	T	A	A	A	T	T	---	---	---	---	---	---	2733	
3619	s	n	Tyr	Leu	ile	Tyr	Lys	Lys	Asp	Phe	ile	Ser	Tyr	Asn	Leu	Tyr	Lys				









RT falciparum.;  
 RL Science 282:1126-1132(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Perlea M., Allen J., Selengut J.J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.  
 RA "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.;  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE001406; AAC71912.2; --  
 DR PIR; C71610; C71610.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2385 AA; 288299 MW; CC7CF642BC7E70B9 CRC64;

Alignment Scores:  
 Pred. No.: 1.09e-05 Length: 2385  
 Score: 184.00 Matches: 236  
 Percent Similarity: 33.20% Conservative: 183  
 Best Local Similarity: 18.70% Mismatches: 362  
 Query Match: 2.39% Indels: 481  
 DB: 5 Gaps: 60

US-09-874-162A-7 (1-4409) x 096216 (1-2385)

QY 688 GAAATGAACAAATTCCTGTTACCTCGTGAAGTCTGTTGGAAGTTTGCCACAAAAA 747  
 DB 69 GUAAPCysAsnLysAsnMetThrIleAsnIleMetLeuLeuGlnAsnGluAsnLys 88  
 QY 748 AGAAGAGTGAAGTGTCCAAATGAAGCAAGTCCACAGTAAAGACAGGTCCTTTG 807  
 DB 89 IleIleLysGluThrCysThrIleLysAsnValValThrAsnLysIleTyHisThrLeu 108  
 QY 808 ATTCTGTACCTCAATCAACAAACCCGGAATTTCCGTCCTCGTTCAGTTTCAGTAAT 867  
 DB 109 PheLeuValIleAsnLysHisTyHisAsnIleLeuCysSerLeuSerPheGluAsn 128  
 QY 868 GAAATTGAACCTAGTAAACAGCCATATGTTGAAGTCTTACTCGTTGCTATTAGAGTGACT 927  
 DB 129 SerPheGluIleLeuAsnThrAsnPheValLysThr----- 140  
 QY 928 CGTCAGGAAGAGAGATTAAATGAATGATTAATGAGAA-----ACCAATGAAAT 981  
 DB 141 -----PheLysGlyLysIleLysSerMetAlaCysThrAsnAsn 154  
 QY 982 ATTGATGTCAATGAAGACTTCCAGCCAGAGAAA-----CGAAATCGTGAG 1029  
 DB 155 IlePheVal-----LeuIleLysLysLysLysLysLysLysLysLysLys 171  
 QY 1030 GATGGGAAAAGACATTTGTTGCAAAATGACATATT-----GATAAAAC 1077  
 DB 172 AsnGlnMetLysSerLysIleLeuAsnGlnAsnValLeuValSerLysHisThrLeuAsp 191  
 QY 1078 AGCGCTTACAGCTTTAGATGGGGAATATGATAGCCATG-----CAGGAATCGAA 1131  
 DB 192 ArgSerLeuLeuLeuMetLysGlyLysValAsnValAsnValIleCysGluSerLysLys 211  
 QY 1132 GAATGTCCAATAGCAAGAAAGACCAATGGGAGACTATTCTTGATGGGAAGAGCGTG 1191  
 DB 212 GluLysLysLysLysLysLys----- 219  
 QY 1192 CCTCCATTGCAACATTTTCTCAGGAGCCTTACAGTTGCGAGTTCACTCTTCGTGGACAGGA 1251  
 DB 220 -----ThrAsp 221

QY 1252 GAGACCAATGATAAATCTACGGCTCCTATT---GCCAACCTCTTGCCACTAGAAATTC 1308  
 DB 222 AsnLysAsnGluLysLysGlyHisMetGluIleLysAspValAsnGluLysLeuAsn 241  
 QY 1309 GAGAGTCTCCATCAGAA-----AACAGCCTGGTTCAGTTAAACCTACTCAAACTATT 1362  
 DB 242 GluLysIleAsnGluLysAsnGluLysIleAsnGluLysAsnGluLysAsn 261  
 QY 1363 GCTGTTAAAGATCATGCTACTACAGATCTACAAACAGAAAGAAAGAGATCTCAAAAT 1422  
 DB 262 GluGluLysAsnGluLysIleAsnGluLysAsnGluLysAsnGluLysLeuAsn 281  
 QY 1423 GAAACCCGCAAAAAATTAGAATATTTATCATGTTCTCTATAACAACAATACAAAGCAA 1482  
 DB 282 GluLysAsnGlu-----AspThrAsnLysAspPro 292  
 QY 1483 CAAACTGAAGCAGAGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1542  
 DB 293 TyrGluGluLysGluAsnAsp-AsnIleProLeuGlyAspHisHisSerValGlnTyrAs 312  
 QY 1543 TATAGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTTATCTCAACTATGTTT-- 1600  
 DB 312 nIlePheThrPheSerIleLeuAsnLysLysGluProAspLeuLysLysIleGlnPheSe 332  
 QY 1601 ----ATCATCCAAAAGGTGCTAGGATAGTGTTCATCAATGAGTGTATGATGGCTCC 1656  
 DB 332 rAsnIleIleLeuProIle-----LysLysMetIleIleCysPr 345  
 QY 1657 TATCGAGAAATCCTCAGCATATTATCGCCCACTGGATTGCTTTAGTCGCAACGGA 1716  
 DB 345 cTyrAspGluLysIleIleIleLeuLeuSerHisLysSerIleValTyrIleThrAs 365  
 QY 1717 CAGATTAAAGAAACACCTATCACATATCTGTGTGCGAGCCCAACAAACGAAACGA 1776  
 DB 365 nLysAsnAsnAspLeuLysAsnMetPheIleIle-----LysG 379  
 QY 1777 AGCATGCTCGAAT----- 1789  
 DB 379 uLeuIlePheAsnSerProIleIleThrThrThrTrpIleAspAsnTyrIlePheLeu 399  
 QY 1790 ----TCTTTGAAT-----CTGAAGATGGGAAG----- 1813  
 DB 399 eTyrPheLeuAsnAsnGluLeuIlePheLeuSerPheAlaLysProCysArgAsnLeuTy 419  
 QY 1814 -----TAGAACAGCAAGAACATATAGTA----- 1837  
 DB 419 rPheTyrLysCysIleAsnAsnTyr-SerHisIleThrSerPhePheTyrLysSerArgAs 439  
 QY 1838 -----GTGCCACAATCGTCTGATTTCCATAGTGATA----- 1870  
 DB 439 nLeuTyrIleSerPheLysThrLysGluIleValCysPheLysIleArgTyrTyrGlu 459  
 QY 1870 ----- 1870  
 DB 459 eProLeuThrValPheLysLysValAspLysAsnLysArgSerTyrIleGlnThrThrG 479  
 QY 1871 -----CCTGCTTACCTCTCCGTCCA 1890  
 DB 479 uGlyAsnTyrIleAspAlaLysTyrLeuPheArgLysArgProArgTyrIleAsn-Thr 499  
 QY 1891 CAAGAAATGGAAGTAGATAGTAGTGAAGATGAAAGGAT----- 1926  
 DB 499 snHisAsnGlnSer-AsnAlaLysAspLysAspGlyAsnAspValIleArgGluGlu 518  
 QY 1927 CTTGAATGGCTTAAGAGAAAAACCATTAACAATTAAGAGATTTCTTGATGTTTAATGAA 1986  
 DB 519 GluAspPheLeuArgAsnAsnAsn-----LysAsnPheSerAspValLysLys 534  
 QY 1987 GGAGAGAAAGAGTATGAAACTCTGGAATCTCCATGTCATGAGCATGGGTGTTTATTGCT 2046  
 DB 535 ArgLysLysArg----- 538  
 QY 2047 GACAATCAATGAATCATGCGCTGTATGCTGTTGTAGAAAATTATGGACAGAAATAATT 2106



DE Hypothetical protein.  
GN PF11\_0206.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_taxid=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Ariguoli S.,  
RA Perte M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.,  
RT "Genome sequence of the human malaria parasite Plasmodium  
falciparum."  
RL Nature 419:498-511(2002).  
DR EMBL; AE014838; AAN35790.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 1749 AA; 208588 MW; 2ABD506375B3EAF9 CRC64;

Alignment Scores:  
Pred. No.: 2,57e-05 Length: 1749  
Score: 179.00 Matches: 202  
Percent Similarity: 37.64% Conservative: 180  
Best Local Similarity: 19.90% Mismatches: 356  
Query Match: 2.33% Indels: 277  
Dbs: 5 Gaps: 54

US-09-874-162a-7 (1-4409) x Q8IIG8 (1-1749)

QY 451 TATAGATTCTTCGAATCTCGAATCTCATAGCAGCAATATTTTGCACAGAACT----- 504  
Db 119 PheLysLeuArgLysLysLysLysAspLysSerLeuPheAspAspAsnThrTyrAsp 138  
QY 505 -----CTTACTTCATGTCATCGAATCTCCAGAACCTCCAGAACACATCAAAAGGAAACATTT 558  
Db 139 MetGluLeuSerTyr---AsnGluLysAspGluLysAspGluArgAsnGluLysAspGlu 157  
QY 559 AAAGTTGATGATATCTTATCAAAAGTAGAGAAA-----ATGAAA 597  
Db 158 Lys---AspGluArgAsnGluLysAspGluLysAspGluArgAsnGluLys 176  
QY 598 GGAGACGAAGATCTCATAGCTTGTGACGCTCATTTGCAGCTTACGTTTCTGTTTCTTC 657  
Db 177 AspGluLysAspGluArgAsn----- 183  
QY 658 CACAAAATGATAGCCATCACCACCAATGACAAAATGACAAAATCTGTTTACCCCTGAA 717  
Db 184 GluLysAspGluLysAspGluArgAsnLysLysAsnGluAsnAsp----- 198  
QY 718 GTCCTGCTGTGAAAGTTGGCCACAAAAGAAAGAGATGTAAGTTGTCCAAATAGGCAA 777  
Db 199 -----GluLysAspLysArgAspLysLys----- 207  
QY 778 GTTCCACAGGTAAAGACAGGTGCTTGTGATTCCTGACCTCAATCAACAAACACCCGGA 837  
Db 208 -----PheAspLysAsnSerPhePheLeuSerProSerValMetGluProGluGlu 225  
QY 838 AATTTCCCGTCCCTT-----GCAGTTTCCAGTAATGAAATTTGAACCTAGTAACAGCCAT 891  
Db 226 GluPheSerTyrIleAsnSerIleAsnAsnGluMetAsnAsnAsnAsn----- 245  
QY 892 ATGGTGAAGTCTTACTCGTTGCTATTAGAGTGACTGCTCCAGAGAGAGAGTTAAT 951  
Db 246 AsnAsnAsnSerSerAsnTyrArgAsnAsnIlePheTyrAspAsnValGluThrPhe 265  
QY 952 GGAATGATTAAATGGAGAAACCAATGAAATATGATGTCATGATGAAGAGCTTCCAGCCAGA 1011  
Db 953 ----- 1011

Db 266 AsnValIleAsnGluThrGlnAsnGluGly-----AsnGluAsnLeuHisIle 282  
QY 1012 AGAAAAAGAAATCGTGAG-----CATGGGAAAAAGACATTTGTTGCACAAATG 1059  
Db 283 LysLysGlyLysValGluAsnTyrPheAsnAspCysLysLysGluPheLeuVal----- 300  
QY 1060 ACAGTATTGTATAAAAACAGCGCTTACAGCTTTAGATGGGAATATGAA----- 1110  
Db 301 -----AspLysAsn-----PheAsnAspGluTyrTyrGluGluAspIle 313  
QY 1111 -----GTAGCCATGCAGAAATGGAAGATGTCCAATAAGCAAGAAAGAAAGCAACA 1161  
Db 314 AsnAsnTyrMetSerAsnAsnLysValGluLysLeuIleAspLysGluLysGluAsn 333  
QY 1162 TGGGAG---ACTATTCTGTATGGGAGAGGCTGCTCCATTCGAAACATTTTCTCAGGGA 1218  
Db 334 AsnGluTyrAsnLysIleAsnGly-----AspAsnPheHisAsp 347  
QY 1219 CCTACGTTGTCAGTTCACCTCTTCGTTGGACAGAGAGACC----- 1257  
Db 348 AsnAsnIleAsnIleThrGluHisLeuAsnAspSerLeuIleTyrSerHisLys 367  
QY 1258 -----AATGATAAATCTACGGCTCTTATTCGCAAAACCTTTCGCCACTAGAAATCA 1308  
Db 368 ThrTyrPheHisAspProAsn-----LeuAsnAspThrAsnSer 380  
QY 1309 GAGAGTCTCCATCAGGAAAAACAGCGCTGTTCAAGTTAAACCTACTCAAACTATTGCTGTT 1368  
Db 381 GlnHisIleAsnAsnGluAsn-----ValIleLeu 390  
QY 1369 AAAGAATCATTCACACAGATCTACAAAAGAAAGAAAGAGATCTCCAAAAGAAAC 1428  
Db 391 LysGluLysGlyGluGluAsnGluSerAsnAsnGlnLysAspTyrAsnLysIle 410  
QY 1429 CGACAAAATTAAGAATA-----TTTATCAGTTTCTCTATAACAAAC 1470  
Db 411 ArgGluSerValSerLeuAsnAsnAsnIleMetAspAsnAsnSerPheLeuTyrAsnAsn 430  
QY 1471 ---AATACAAAGCAACAACTGAAGACAGATGACCTGCATTCGCTTGGTGTACTCTG 1527  
Db 431 LysAsnAspArgMetSerCysGluTyrAsnGluAspArgAsnIle-----CysSerGlu 448  
QY 1528 AACTGCCGCAAACTTTATAGT-----TTACTCAAGCATCTTAACCTC----- 1569  
Db 449 AsnLeuLysAspIlePheAsnAsnGluGlnGluLysGluTyrMetLysMetAsnIleHis 468  
QY 1570 -----TGCATAGCAGATTTATCTCAACTATGTT-----TATCATCCAAAA 1611  
Db 469 AsnAsnAsnAsnAsnAsnLysTyrIleTyrAspAsnIleAsnLeuTyrHis---Lys 487  
QY 1612 GTGCTAGCATAGATGTTTCTATCAATGAGTGTATGATGGCTCC----- 1656  
Db 488 GlyAspAsnValHisMetAsnValAlaThrIleLysAspHisAsnAsnValAspVal 507  
QY 1657 -----TATCCAGGAATCCT-----CAGGATATTCAT----- 1683  
Db 508 IleSerLysLysAspHisIleAsnIleProSerValHisLysAsnValHisValSerPhe 527  
QY 1684 -----CGCCAACTCGGATTTGCTTTTGTAGTCGCAACCGGACCATTAAGAGAACACT 1734  
Db 528 CysPheGlyAsnAsnGlyThrPheTyrSerLysAsnSerLysIleLysTyrGlnSer 547  
QY 1735 ATCACACATATTCTGTGTGTCAGGCCAAACCAACAAAGCAAGCATGCTCTGAATTTCTT 1794  
Db 548 LeuIleAsnValIleGluIleHisIleAsnLysLysArgLysMetHisIlePheAsn 567  
QY 1795 GAATCTGAATGAGGGAAGTAGAACACAAAGACATATAGTAGTGGCCACCAATCGTCTG 1854  
Db 568 GluLysAsnAsnGluGlu-----AsnAsnIle 576  
QY 1855 TATTTCATAGTACATACCTGCTTACTCTCCGTCACAAAGAAATGGAAGTAGATAGTGA 1914  
Db 577 TyrMetAsnSerAsn---CysAsnTyrAsnAsnTyrCysAsnArgAsnValAspSerAsn 595

QY 1915 GATGAAGAGTCTGAAATGGCTAAGA---GAAAAACCATTTACACAAATGGAAGAGTTT 1971  
 Db |||||  
 QY 596 AspVallyHisCysTyrHisAsnMetGluLysPheIleTyrCysIleLysAsnBhe 615  
 Db |||||  
 QY 1972 -----TCTGATGTTTAATGAAGAGAGAAAGATGATGAA 2007  
 Db |||||  
 QY 616 ProGlyProPheSerArgLysSerAsnLysValAspHisIleGluGluPheLeuLys 635  
 Db |||||  
 QY 2008 CTCTGGAATCTCATGTCATGAAGCATGGTTTATGTCGACAAATCAATGAATCATGCC 2067  
 Db |||||  
 QY 636 GlyTyr---IleAsnIleAsnLysSerArgSerValAsnAspAsn----- 649  
 Db |||||  
 QY 2068 TGTATGCTGCTGTTGTAGAAATATTGACACAGAAATTAATTAAGAAAGATTTATGTCGAAC 2127  
 Db |||||  
 QY 650 -----IlePheGluAsp-----IleGlnLysAsnCysLeuTyrAsn 661  
 Db |||||  
 QY 2128 TTATGCTTCATCTAGTCAGC-----ATGCATGACTTTTAATCTTATTAGCATAATG 2178  
 Db |||||  
 QY 662 PheLeuLeuAsnIleLeuLysLysProTyrLeuThrAsnPheAsnLeuLysGluIle--- 680  
 Db |||||  
 QY 2179 TCAATAGATAAGCTGTTACCAAGCTCCGCGTAATGCGACAAA----- 2223  
 Db |||||  
 QY 681 SerIleGluLysLysAsnAspLeuAsnLysAsnMetAsnHisLysGlyAsnAsnIleIle 700  
 Db |||||  
 QY 2224 -----TTAGAAAAGGGGGAATCTGCTTCCCTGCAACGAGAAATTAACCTGAAGAA 2274  
 Db |||||  
 QY 701 SerTyrPheLeuAspAsnTyrGlnGluLysLysAspIleLysAspAspIleTyr----- 718  
 Db |||||  
 QY 2275 CAAATGGGACAGCAATGGATTAGTGAATTAATCACTCAAGAGAAAGCTTTGGAACA 2334  
 Db |||||  
 QY 719 -----LeuSerAspAspCysAsnGluAspGluAspIleAspIle 731  
 Db |||||  
 QY 2335 GATAGTGTCTCAGG-----GTTTCAAAACAG 2361  
 Db |||||  
 QY 732 AspSerLeuSerAsnAsnGluMetAspGluPheAspIleAsnGluTyrValLysAsnAsp 751  
 Db |||||  
 QY 2362 AGCAAAAACAAAACTGAAAAGCTCTAACCCTGTTATGACCAACAACTGAAAT-- 2419  
 Db |||||  
 QY 752 SerAsnLysMetLys-----LysTyrAspGluProCys-AsnAsnAsnHisIleAsnIle 769  
 Db |||||  
 QY 2420 -TACATTTTGGGAATTCATCTCTCAAGAAATATGTTTCTTTTAAATCATATGTTCCA 2478  
 Db |||||  
 QY 769 eTyAsnProThrAsnAsnAsnSerAsnAsnIleMetAsnAsnAsn-ValAspLysL 789  
 Db |||||  
 QY 2479 AACAGGCACTGTATAGT-----GAAGTAAATGATTTCAACA 2514  
 Db |||||  
 QY 789 yLysAsnValLeuAsnLysAspGlnLysLysCysTyrThrAspValTyrGluAsnSerL 809  
 Db |||||  
 QY 2515 AG-----GATATTTGTATCAGGGTCTACTTCACTTCATTCATTCAGCATTTACATGT 2565  
 Db |||||  
 QY 809 yAsnIleAsnAspIleCysIleAsnAsnAsnThrTyrPhe----- 822  
 Db |||||  
 QY 2566 ATATCACCTTTTATGATGATCAATTAACATCTCTACTTTAAGCATGAAAGCAATATTT 2625  
 Db |||||  
 QY 823 -----PheAspPheIleAsn-----LysGluPheIleM 833  
 Db |||||  
 QY 2626 CAAAGATATTTAACTCAACAAATGTCATCAAAATATGTTGAATGATCTAGAAATATT 2685  
 Db |||||  
 QY 833 etGluLeuTyrLysLeuGluAsnLysAsnGluArgIleThrHis---AspAspIleTyrL 852  
 Db |||||  
 QY 2686 TCATATAT-----AAATCAGAAATTTTGTGATTTATG 2718  
 Db |||||  
 QY 852 euLeuTyrPheHisMetCysIleTyrAsnSerTyrLysAlaThrAsnValCysIleLysL 872  
 Db |||||  
 QY 2719 AACGGCTGTTTCTTACTTCTTAATTTGTGACATTTTCTTGGGAGGAAATTTGGAAT 2778  
 Db |||||  
 QY 872 yGluLeuTyrLysTyrPhePheLeuValLeuArg-----Lys-TyrAsn 886  
 Db |||||  
 QY 2779 GGTCCCTTTTAAAGATG-----AAGTGGTCTTCATAT-----GTCAAC 2820  
 Db |||||  
 QY 887 LysLysLysTyrTyrLysMetLeuAspLysTyrIleSerTyrIleLysArgSerIleAsn 906  
 Db |||||

QY 2821 TACAGAAAGGAAAAAATAGAAATTCAGAGATTTTATGAAATATATATGCAATTACT-- 2878  
 Db |||||  
 QY 907 AsnGluLeuGluMetLysAsnAspMetLysAsnIleTyr-LysIleTyrAsnTyrAsnIle 926  
 Db |||||  
 QY 2879 -----ATTTCAGTCAAACTTTGATCCCTGTTTGTGAAATCATTTGTCAATTCGG 2928  
 Db |||||  
 QY 926 eTyAspGluIleCysThrGluAlaPheValPhe-----PheLeuCysIleLeuAs 943  
 Db |||||  
 QY 2929 AATGAAAAATAT-----AATGTAATTTTACATCATATAGTTCCCTTTTACAA 2976  
 Db |||||  
 QY 943 nLysLysAsnMetValPheLysLysAsnIleLeuValAsnTyrTrpTyrAlaPheTyrTh 963  
 Db |||||  
 QY 2977 TTAATAATAGCACTTCTCATCTTATGCTGTTGAGAGATATTAATAATTTTACATG 3036  
 Db |||||  
 QY 963 rLeuIlePhe-HisAsnPheIlePheGlnTyrLysGluAsnGluIleGlnPheAlaAspT 983  
 Db |||||  
 QY 3037 TTGACAGTGAATGCTATGTTGTTTATTAAGATTA 3071  
 Db |||||  
 QY 983 yLysGluAspIleIleAsnPhePheIleTyrLeu 994  
 Db |||||  
 RESULT 9  
 O59904 ID O59904 PRELIMINARY; PRT; 1314 AA.  
 AC O59904;  
 DT 01-AUG-1998 (T-EMBLrel. 07, Created)  
 DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Silent information regulator 4.  
 GN SIR4.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98198828; PubMed=9539421;  
 RX Astrom S.U., Rine J.;  
 RT "Theme and variation among silencing proteins in Saccharomyces  
 cerevisiae and Kluyveromyces lactis."  
 RL Genetics 148:1021-1029(1998).  
 DR EMBL; AF035007; AAC39438.1; -.  
 DR PIR; T09481; T09481.  
 SQ SEQUENCE 1314 AA; 146735 MW; 3C78CBB63319F8A6 CRC64;  
 Alignment Scores:  
 Pred. No.: 2,69e-05 Length: 1314  
 Score: 178.50 Matches: 188  
 Percent Similarity: 35.87% Conservative: 147  
 Best Local Similarity: 20.13% Mismatches: 360  
 Query Match: 2.32% Indels: 240  
 DB: 3 Gaps: 43  
 US-09-874-162A-7 (1-4409) x O59904 (1-1314)  
 QY 59 GCATCGCCGCCCTCTCTTCTTCCAAATPACCTGCGGATTCGGGGCTCGGACTCCACT 118  
 Db |||||  
 QY 110 SerSerArgGlnProSerProSerProSerProAlaAlaSerGlySerGlnTyrGlnLys 129  
 Db |||||  
 QY 119 TCCCCACCTCGCCGACC---TCATCGACCATCGAGGACAAAC----- 159  
 Db |||||  
 QY 130 GlnLysSerTyrThrAsnArgLysSerSerThrSerHisThrGluMetLeuLysSerLeu 149  
 Db |||||  
 QY 160 -----CACATCGATACATACCA--- 177  
 Db |||||  
 QY 150 TyrValGlyAspSerLysValSerSerProProGlnSerAsnIleAspLysAlaProGln 169  
 Db |||||  
 QY 178 -----CGGTTTTAGAAAAACAAAGATTA--- 201  
 Db |||||  
 QY 170 ArgLysProLeuGluArgThrGlyLeuSerLysArgValValGluAspGluValLeuArg 189  
 Db |||||  
 QY 202 ---CAGCGCCCAACTATGTTGCTCCCTGAGTTACATAAATAGATTTCATGACAGATCTGCC 258  
 Db |||||  
 QY 190 ArgGlnGlnSerSerSerSerSerGlyArgIleHisLysHisIleGluSerProPhe 209  
 Db |||||



```

Db      854  ---GluAspLeuAsnLysThrGluValIleGlnAsnLeuThrSerGluIlelle---Gln 871
QY      2152 CATGACTTTAATCTTATTAGCATATGTCATAGATAAAGCTGCTTACCAAGCTCCGTGAA 2211
Db      872 SerGluMetLysIleAlaSerLeuValGlyIleAsnHis-----GlnLeuArgGlu 888
QY      2212 ATGCAGCAAAATAGAAAG-----GGGAATCTGCTTCCCTGCTCAACAGAGAAATA 2265
Db      889 LysLeuGluGluLeuGluLysIleGlnAlaLysLeuValGlnGluLeuAspThrValLeu 908
QY      2266 ACTGAAGAACAAAT-----GGGACAGCAAAATGGATTAGTGAATAAATCACTCA 2313
Db      909 LeuGluSerLysGlySerPheSerAlaGlyThrAlaLysGlu---AlaGluValAsnAsn 927
QY      2314 AAAGAGAAAGCTTGGAAACAGATAGTGTCTCAGGGTTCACAAACAGAGCAAAACAA 2373
Db      928 LysAspLys-----ThrGluSerIleAsn-----AsnGluGlnAlaArgGlu--- 941
QY      2374 AAATCTGAAAAGCTCTAACCCCATGTTATGGACAAACACTG 2415
Db      942 -----LysSerLeuThrAlaAlaHisGlyLysLeulle 952

RESULT 10
Q8IBL9 ID Q8IBL9 PRELIMINARY; PRT; 1422 AA.
AC Q8IBL9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF07_0097.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
KW EMBL; AL844506; CAD50981.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1422 AA; 167923 MW; 0910F885329A8696 CRC64;

Alignment Scores:
Pred. No.:
Score: 3.276-05 Length: 1422
Matches: 174
Percent Similarity: 33.30% Conservative: 133
Best Local Similarity: 18.87% Mismatches: 219
Query Match: 2.31% Indels: 297
DB: 5 Gaps: 43

US-09-874-162A-7 (1-4409) x Q8IBL9 (1-1422)
QY 65 CCGCGCCTCTCTCTCTCCA----- 85
Db 623 ProProProSerThrAspAsnArgPheAsnIleSerArgLysSerLeuLeu 642
QY 86 ATACTCGCATTCGGGGCTCGGAGCTCCACTTCCCAACCTCGCCGACCTCATCGAGC 145
Db 643 IleProAsnAspLysGlnMetLysIleSerAsn-----MetSerAsnThr 657
QY 146 ACATCGAGGAC-----AACCATCGAT-----ACAGATCCACCGGTTTGAAGAAACAA 195
Db 658 ThrGlyMetProTyrAsnHisValAsnMetMetAsnAsnProAsnMetIleAsnArgGln 677
QY 196 GAATTACAGCAGCAACCATGTTGCTGCTGAGTTACATAATATAGATTCAATGACAGATGCT 255
Db 678 Ser-----HisValSerLeuLeuAsnAsnThrPheThrGluPhe 690
QY 256 CCCGCCGAGAGCAG-----GATGCCCTAAAGAAAGATTCAGCCGAGAGCTCTCGCTG 309
Db 691 GluMetLysAspLysAsnIleGluAsnSerAsnLeuAsnAlaProAsnIleAsnIle 710

```

```

QY 310 ACTCTGTCTCAGCTCA-----GTGCTCTCGAGGAATGTGTCC 345
Db 711 ThrSerAsnAsnAsnIleProProProAspThrAspIleAsnLysAsnLysMetIle 730
QY 346 ACTCCCCCAGCCACAGCAGTGGAGCTTACTCCCCCGTGACCCACCCATCCACCCC 405
Db 731 AlaProProProTyr---AlaGlyAspAsnThrAlaSerValAsnProHisIleProAsn 749
QY 406 TCCTCTTCA-----TTCCGACGACGACTCCGACAGACCAACACACATCTAT 453
Db 750 PheHisSerGlyValMetAsnArgAsnSerIleSerProIleTyrProThrSerIle--- 768
QY 454 AGATTCTTCCGAATCTCAGTACACCAATATTTTTCGACAGAACTCTTACTTAC 513
Db 769 -----ArgSerAsnAsnMetIleAsnAsnMetMetAsnAlaAspGlyThrVal 785
QY 514 -----ATGCTCTCAATCGAAATCCAGAACAAATCAATCAATCAATCAATCAATCAAT 561
Db 786 LeuProAlaAsnMetAsnHisProAsnAsnMetLeuAsnVal----- 799
QY 562 GTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGCAAGAAATCTCATAGTTG 621
Db 800 ThrAsnAsnMetMetValSerProHisProMetAsnAsnGluMetLysAsnMetSerPhe 819
QY 622 TCAGCTCATTTGCAGCTTACGTTTACTGTGTTTCTTCCACAAAATGATAAGCCATCCA 681
Db 820 AsnLysAsnMetAsnIleSerLysThr-----Pro 829
QY 682 AACTCAGAAAATGAACAAAAATCTGTTCACCTCGAAGTCTCTGTGTGAAAGTTTGCCAC 741
Db 830 Asn-----AsnAsnValIleLeu----- 835
QY 742 AAAAAAAGAAAGATGTAAGTTGTCTCAATCAATCAATCAATCAATCAATCAATCAAT 792
Db 836 -----ThrLysAsnValAsnSerLeuMetAsnIleGlnArgGluProIleAsnSer 853
QY 793 AAGCAGGTGCTTTG---ATTCTCGACCTCAATCAACAAACAAACCCGAAATTTCCCG--- 846
Db 854 TyrAsnAsnProLeuGlyPheLysProLeuAsnTyrAspAspProAsnSerIleProIle 873
QY 847 -----TCCCTTGCAGTTTCCAGTAATGAATTT 873
Db 874 AsnTyrAspGluGlyAlaTrpPheAsnAsnAsnGlnMetAsnGlyThrAsnAsnAsnMet 893
QY 874 GAACCTAGTAACAGCCATATGTTGTAAGTCTTACTCTGTTCTATTAGAGTACTCTCCA 933
Db 894 TyrHisMetAsnAspAsnIleLeuLysProAsnGluAspLeuIleArgIleSer----- 911
QY 934 GGAAGAGAGAGATTAAATGGAATGATTAAT----- 963
Db 912 GlyIleLysAspAsnAsnAsnAsnAsnAsnAsnGluGluGluLysLysLysLys 931
QY 964 -----GGAGAAACCAATGAAATATTTGATGTCATGTAAGAGCTTCCAGCCAGAGAAA 1017
Db 932 GluLysGlyGluGluThrGluGlyGluGluValAspGluGln-----GlyGluIleGlu 949
QY 1018 CGAAATCTGTGAGGATCGGGAAGACATTTGTT-----GCACAATGACAGTATTGAT 1071
Db 950 ArgAsnGlyGluGluAsnGluLysLysLysLysAspAsnLysAsnMetLeuIleGlnThr 969
QY 1072 AAAAAAGCGCTTACAGCTTTTATAGTGGGGAATATGAAGTAGCCATGAGGAATGAA 1131
Db 970 HisHisHisLysAspGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 989
QY 1132 GAATGTCCAAATAGCAGAAAGAGCAACATGGAGACTATTCTTGATGGGAGAGGCTG 1191
Db 990 GluLeuThrLysGluGluLysLeuHisGluTrp----- 1000
QY 1192 CCTCCATTCGAAACATTTTCTCAGGACCTAGTGTGACGTTTCTGTTTCTGTTGACAGGA 1251
Db 1001 -----LeuGluTyrGluLeu-----ThrAsn 1007

```











```
QY 1288 CCTCTTGCACATAGAAATTCAGAGAGCTCCATCAGGAAAAACAACGCTGGT----- 1338
Db      |||      |||      |||      |||      |||      |||      |||      |||
656 ProThrThrIleThrAlaSerAlaThrAlaGluGluLysGluAspSerGluThrIle 675
QY 1339 -----TCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCATTGACTACAGAT 1389
Db      |||      |||      |||      |||      |||      |||      |||      |||
676 LeuGluGluSerIleLysProLysAspHisIleMetGluGluThrAspSerLysThrAsp 695
QY 1390 -----CTACAAACAAGAAAAAGAAAG 1410
Db      |||      |||      |||      |||      |||      |||      |||      |||
696 AsnCysAsnAspIleSerMetGluAspAlaAsnSerAsnAsnAsnSerAsnAsnAsn 715
QY 1411 GATACCTCCAAATGAAACCGACAAATTAAGAAATTTATACAGTTCTCTATACAAAC 1470
Db      |||      |||      |||      |||      |||      |||      |||      |||
716 AspAsnAspAsnGluAsnAsnAsn-----AsnAsnAsn 727
QY 1471 AATACAAAGCAACAACTGAAGCAAGAGATGACCTGCATTGGCTTGGTACTCTGAAC 1530
Db      |||      |||      |||      |||      |||      |||      |||      |||
728 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn----- 740
QY 1531 TGCCGCAACTTATAGTTTACTCAAGCATCTTAACCTCGCATAGCAGATTATCTTC 1590
Db      |||      |||      |||      |||      |||      |||      |||      |||
740 ----- 740
QY 1591 AACTATGTTTATCATCCAAAAGGTCTAGATAGATGTTCTTATCAATGAGTGTATGAT 1650
Db      |||      |||      |||      |||      |||      |||      |||      |||
741 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 760
QY 1651 GCTCCTATGAGGAAATCCTCAGGATATTCATGCCAACCTGGATGTTCTTTAGTCGC 1710
Db      |||      |||      |||      |||      |||      |||      |||      |||
761 AsnGluLysSerGluAsnGluAsnGluInSerLysGlnAspThrLysSerGluSerAsn 780
QY 1711 AACGACACGTTAAGAGAACCTATCACACATATCTTGTTGTGAGGCCAAACGACACA 1770
Db      |||      |||      |||      |||      |||      |||      |||      |||
781 AsnGlyThrAsnLysMetAsnGlyValAspHisSerPheIleAspAsnAsnGluLysAsp 800
QY 1771 AAAGCAAGCATCTGAA-----TTTCTTGAATCTGAAGATGAGGAAAGTAGAACAG 1821
Db      |||      |||      |||      |||      |||      |||      |||      |||
801 AsnArgLysLeuThrGluGluGluValAlaLeuGluAsnThrGluGlnAspIle----- 818
QY 1822 CAAGAACAATATAGTAGTGGCCCAATCGTCTGTATTTCCATAGTAGTACCTGTTACCT 1881
Db      |||      |||      |||      |||      |||      |||      |||      |||
819 -----AsnIleCysAspAsn 823
QY 1882 CTCCTGCACAGAATGGAATGATAGTAGAAGATGAAAGGATCCCTGAATGCTGAAGA 1941
Db      |||      |||      |||      |||      |||      |||      |||      |||
824 GlnProProGlnGluLysGluGlnGluLysGluGluGluLysGluGluGluLys 843
QY 1942 GAAAAACCATTACAAATTAAGAGATTTCTGTGATTAATGAAGAGAGAAAGAGTG 2001
Db      |||      |||      |||      |||      |||      |||      |||      |||
844 GluLys-----GluGlu-----GluLysGluGluLysLysLysGluVal 856
QY 2002 ATGAACCTCTGGAATCTCATGTCATGAGCATGGTGTATTGCTGACAAATCAATGAAT 2061
Db      |||      |||      |||      |||      |||      |||      |||      |||
857 GlnLysGluGlyGluIleLeuAsnValLysGlu-----GluAsnThrGluGlu 872
QY 2062 CATGCTGTATGCTGTTGTAGAAAATTTATGACAGAAAAATAATTAAGAAATTTATGT 2121
Db      |||      |||      |||      |||      |||      |||      |||      |||
873 LysThrLeuGlnLeuGluIleLysLysIleGlnGluLysValSerLysLysMet--- 891
QY 2122 CGAAACTTCATGCTCATCTAGTCAGCATGATGACTTATCTTATTAGCATATGTCA 2181
Db      |||      |||      |||      |||      |||      |||      |||      |||
892 GluThrIleGlnGluMetValGlu---LysAspPheGluAspIle----- 906
QY 2182 ATAGATAAAGCTGTTACCAAGCTCCGTGAAATGCAGCAAAAATTAGAAAAGGGGAATCT 2241
Db      |||      |||      |||      |||      |||      |||      |||      |||
907 IleGluLysGluGluLysGlnIleLeuThrThrGlnGlnValSerLysGluAsnGluAsn 926
QY 2242 GCTTCCCTGCNAACGAGAAATAACTGAAGA----- 2274
Db      |||      |||      |||      |||      |||      |||      |||      |||
927 GluAsn---GluAsnGluLysValValGluSerIleIleIleLysLysGluIleGluIle 945
```

```
QY 2275 CAAATGGGACAGCAAAATGGAATTTAGTGAATTAACATAAAAGAGAAAGCTTTGGAAACA 2334
Db      |||      |||      |||      |||      |||      |||      |||      |||
946 GluAsnGluAsnGluAsnGluAsnLysAsnLysAsnGluThrThrPheGluLys 965
QY 2335 GATAGTGTCTCAGGGGTTTCAAAACAGACGCAAAAACAAA 2376
Db      |||      |||      |||      |||      |||      |||      |||      |||
966 ValAspIleAsnLysGluAsnGluGluSerLysGluAsnLys 979

RESULT 13
Q8IM18
ID Q8IM18 PRELIMINARY; PRT; 2033 AA.
AC Q8IM18;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0073.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RA "Genome sequence of the human malaria parasite Plasmodium
RA falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014817; AAN36685.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2033 AA; 241130 MW; AA7F62DFF694870 CRC64;

Alignment Scores:
Pred. No.: 7,85e-05 Length: 2033
Score: 173.00 Matches: 165
Percent Similarity: 34.78% Conservative: 171
Best Local Similarity: 17.08% Mismatches: 333
Query Match: 2.25% Indels: 297
DB: 5 Gaps: 36

US-09-874-162A-7 (1-4409) x Q8IM18 (1-2033)
QY 154 GACAACCATCATCATACAGATCCACGGGTTTAGAAAAACAAGATTACAGCAGCAACC 213
Db      |||      |||      |||      |||      |||      |||      |||      |||
654 GluAsnAsnIlePheAspLysSerMetIleThrGlnLysLysGluLysLysLysAsn 673
QY 214 TATGTCCTCGATGATTACATAAATGATTATCATGACAGATGCTCCCGCAGAGAGGAG 273
Db      |||      |||      |||      |||      |||      |||      |||      |||
674 PheIleGluIleGluThrAsnIleGluAsnAsnHisLeuAsnAsnGluMethIleLysThrAspAsp 693
QY 274 TCCTTAAGAAAGAAAGATTACCGAAGCTCTCGCTGACTCTGCTCCAGCTCAGTCTCGA 333
Db      |||      |||      |||      |||      |||      |||      |||      |||
694 AsnThrGluThrLysLysLysLysGlnIleSerLeuLeuLysSerAsnIleLysHis 713
QY 334 GGAATGTGTCCACTCCCCACGCCACAGCAGGTGGAAGCTTACTCCCCCGTACCCCA 393
Db      |||      |||      |||      |||      |||      |||      |||      |||
714 AsnAsnLeuArgIleHisLysAsnThrAspLysGlyAsnAsnLysLysTrpIleSerGly 733
QY 394 CCATCACCCCTCTCTTCATTCGCCGACGACACTCCGACAGCAACACAGATCTAT 453
Db      |||      |||      |||      |||      |||      |||      |||      |||
734 AsnIleIle---AsnAsnGluIleThrThrLysAspIleThrTyrAsnAsnLeuTyr 752
QY 454 AGATTCTTCGAACTCGGAATCTCATAGCACCACCAATATTTTGCACAGAACTCTTACTTAC 513
Db      |||      |||      |||      |||      |||      |||      |||      |||
753 LysIleAsnArgValLys-----LeuGluLys 761
```

Qy	514	ATGTCTCATCGAAACTCCAGAACAAACATCAAAAGGAAACATTTAAA-----	561
Db	762	GluAspAsnAsnAsnGlyArgIleArgPheLysArgGlyThrPheAsnGluThrTyrAsp	781
Qy	562	-----GTTGATGATATGTTATCAAAAGTA	585
Db	782	AsnIleTyrHisHisLysLysCysTyrAspLeuValLeuAspGluLeuLeuLysValThr	801
Qy	586	GAGAAATGAAAGGAGAGCAAGMATCTCATAGCTTGTGACGCTCATTTGCAGCTTACGTT	645
Db	802	ArgLysLysLysAsnLysLysHisTyrHisGluTyrAspLys-----	815
Qy	646	ACTGGTTTCTCCAAAAAATGATAG-----	690
Db	816	-----MetHisAsnHisGluLysAsnIleLeuThrAsnAsnSerAsnAsnAsnAsn	832
Qy	691	AATCAACAATACTCTGTACCTCGAAGTCTGCTTGTGAAAGTTTCCACAAAAAAGA	750
Db	833	AsnAsnHisAsnAsnAsnArgPheAspIleAsnLysMetLysGluProAspGluLysArg	852
Qy	751	AAGGAT-----GTAAGTTGTCCAATAAGGCAAGTTCACCAGGTAAAAAGCAG	798
Db	853	LysAsnAspLysSerSerIleThrCys-----	861
Qy	799	GTGCTTTGATCTCTGACCTCAATCAAAACAAACCCGGAATTTCCCGTCCCTTGCAATT	858
Db	861	-----	861
Qy	859	TCCAGTAATGAATTTGAACCTAGTAACAGCCATATGGTGAAGTCTTACTCGTGTCTATTT	918
Db	861	-----	861
Qy	919	AGAGTACTCGTCCGGAAGAAGAGAGTTTAATGGAATGATTAATGAGAAACCAATGAA	978
Db	862	-----AspAsnGlu	864
Qy	979	AATATTGATGTCATGAAGACTTCCAGCCAGAGAAAA--CGAATCTGAGGATGGG	1035
Db	865	AsnIleArgIleAspTyrLysArgSerAlaMetGluThrIleArgAsnGlnLys-----	882
Qy	1036	GAATAAGACATTTGTTGCACAAATCACAGTATTTGATAAAAAACAGCGCTTACAGCTTTTA	1095
Db	883	-----IleHisGlnValThr-----	891
Qy	1096	GATGGGAATATGAAGTAGCATGCAGAAATGGAGAANTGCCAATAGCAGAAAGA	1155
Db	892	ValThrLysTyrGluSerLysLeuIleGluLeuGluLysMetIleLysLeuLysGlu	911
Qy	1156	GCAACATGGGAGACTATTCTTGATGGGAAGAGGTGCTCCATTTCGAAACATTTTCTCAG	1215
Db	912	LeuLeuLysLysGluAspIleGluAsnGluGlnLeuLysSerPheLeuIleLeuGlnAsp	931
Qy	1216	GGACCT-----ACGTTGCAGTTCATCTCTGTTGGCAGGAGAGCAATGATAAATCT	1269
Db	932	GluGluIleAsnAlaLeuLysGluGluLeuLysIleLeuSerGluGlnAsnAsnLysLys	951
Qy	1270	ACGGCTCTATTGCCAAACCTCTTGCCACTAGAAATTCAGAGATCTCCATCAGGAAAC	1329
Db	952	-----LysLysIleSerLysThrLysLysAspLysSerHisGlnAsnThr	966
Qy	1330	AAGCTCGTTTCAGTTAAACCTACTCAAACTA-----	1368
Db	966	rThrPheTyrLys--AsnLeuIleLysSerLysProSerLysPheMetGluIleIleLe	985
Qy	1369	AAAGAATCATGTACAGATCTACAACAAGAAAGAAAGGATCTCCAATGAAAAC	1428
Db	985	uLysAsn-----ValLysGluLysLysGlnGluLysThrCysAsnPheLysGluLe	1003
Qy	1429	CGACAAAAATTTAGAATAATTTATCATGCTTCTCTATAACAACAATAACAGCAACAACT	1488
Db	1003	uAspLysAsnAsnAspGluMetIle-MetLys-----	1013

1489	GAAGCAAGAGATGACCTGCANTCCCTTGGTGATCTACTGAACTGCCGCCAAACTTTATAGT	1548
1014	-----AspProIleIlePheAsp-AsnLysAsnAsnIleAsnPro	1026
1549	TTACTCAAGCATCTTAAACTCTGCCATAGCAGATTTATCTTCAACTATGTTTATCATCCA	1608
1027	IleValLysAsnValAsn-	1032
1609	AAAGGTGCTAGGATAGATGTTTCTATCAATGAGTGTATGATGCCTCTCTATGCAGGAAAT	1668
1033	-----SerLeuSerThrSerCysAsnSerValHisSerCysTyrProLysAsp	1049
1669	CCTCAGATATTCATCGCCAACTGGATTTGCTTTAGTCGCCAACGGACCAAGTAAAGAGA	1728
1050	LysLysAsnIleGluAsnGluGluAsnGluLysAsnGluLysAsn-	1064
1729	ACACCTATCACATATTTCTGTGTGTCAGGCCAAAACAGCAAAAGCAACATGTTCTGTAA	1788
1065	-----GluLysAsnGluLysAsnGluLysAsnGlu-	1074
1789	TTTCTTGAATCTGAAGATGGGAAGTAGAACACAGCAAAAGACATATAGTAGTGGCCCAAT	1848
1075	-----LysAsnGluGluAsnGluLysAsnGlnLysGln-	1090
1849	CGTCTGATTTCATGTATGATACCTGTTACCTTCCTCGTCCACAGAAAATGGAAGTAGAT	1908
1091	GluThrAlaIlePheAsnAspMetProProAsnIleLysAsnAsnGlnIleAspMetAsn	1110
1909	AGTGAA-----GATGAAAAGATCCTGAATGGCTAAAGAGAAAACCAACC	1950
1111	AsnThrPheSerTyrlleCysCysAspMetArgLysArgArgSerLeuAspSerSerThr	1130
1951	ATTACAAATTCAGACAGTTTCTGATGTTTAATGAAGAGAGAAA-----	1995
1131	GlyPheIleIleLysGluLysAsnAspGluAsnAlaSerHisLysLysLysThrMetPhe	1150
1996	-----GAAGTGATGAAAACCTCTGG	2013
1151	CysProMetTyrlLysAsnGluTrpArgIleAsnLysAsnAspGluMetArgLysGluTyrl	1170
2014	AATCTCCATGTCATGAGCATCGGTTTATGCTGACAAATCAATGAATCATCATGCTGTATG	2073
1171	HisProHisGluThrSerAsnGlyGlnIle--HisAsnTyrlValThrAsnLysTyrlPro	1189
2074	CTGTTTGTAAGAAATATGGACAGAAAATAATTAAGAAGAAATTTATGTGCAAACTTTCATG	2133
1190	IleLysAsnAspAsnThrLysIleAsnLeuLysValAsnIleSerGluGlyLysLys	1209
2134	CTTCATCTAGTCATGCATGATCTTTAATCTTATTAGCATATAGTCAATAGATAAAGCT	2193
1210	GluArgIleAlaSerVal--AsnPheAsnLysCysGlu-----AspAspAsn	1224
2194	GTTACCAAGCTCGTGAAATGACGACAAAATTAGAAAAGGGCAATCTGCTTCCCTGCA	2253
1225	LysAsnAspIleGluAsnMetAsnGluArg-----ThrSerAsn-IleGln	1239
2254	AACCAAGAAATACTGAAGAACAAATATGGGACAGCAAAATGGATTTAGTGAATTAATCA	2313
1239	nThrMetAsnLysIleGluAsnGlnIleLysAsnLysIleAspCysAsnAsnMetGluAs	1259
2314	AAAGAGAAAGCTTTGGAACAAGATA--GTTGCTCAGGGGTTTCAAAAC-----	2359
1259	nLeuHisLysTyrlAsnLysGluSerAsnThrAsnAspAlaTyrlHisAsnLeuAsnLeuCy	1279
2360	-----AGACAAAAAACAAAAAC-----TC	2379
1279	sProSerThrIlePheArgLysLysAsnAsnAsnLysLysAsnAsnValLysIleCysGln	1299
2380	TGAAGAGCTTAACCCCATGTTATGGACAACACTGAAATTTACATTTTATGGGAATCA--	2437
1299	nGluLysIleAsnProHisIleVal-----HisIleAsnTyrlIleSerAsnAspAsnPh	1317
2438	-----TCCTCTAAGAAATATGTTTGTGT-----	2461

```

Db      1317 eGlyGlnMetArgLysSerGlyLysAsnLysTyrPheLeuArgLysValLysGluGluLys 1337
QY      2462 -----
Db      1337 sLeAsnGluGluLysLysGluAsnIlePheSerLysPhePheSerTyrIle-LysLysL 1357
QY      2482 AGGCACTGTAGATGAAGTAATGATTTCAACAAGCATATT---TGTATCAGGGTTCTAC 2538
Db      1357 ysSerileCysAsnGluLysLysAspHisAsnLysAsnLeuArgTyrAsnSerHisV 1377
QY      2539 TTCACCTTCATTATGACGATTACATGATAT-----ATCACTT 2574
Db      1377 aHisLeuArgThrLysAsnTyrIleTyrLysSerProAspAsnLysArgSerGlnSerP 1397
QY      2575 TTATTGATTCATTAAACATCTGTACTTTAAGCATGAAAGCAATATTTCAAAGTATT 2634
Db      1397 hEileAsnLeuPheLysAsn---CysThr-----AlaGlnAsnGlySerLysAsnT 1413
QY      2635 TTAAACTCAACAATGTGCATCAATATGTTGAATGATCTAGAATATTTCATATATA 2694
Db      1413 yThrAspTyrValLeuThrAspAsnAsnPheAsnSerileAsnThrTyrGlnIleAsnA 1433
QY      2695 AATCAGAA 2702
Db      1433 sPLysGlu 1435

RESULT 14
Q8IKP5 PRELIMINARY; PRT; 1163 AA.
AC Q8IKP5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF14_0559.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.,
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
DR EMBL; AB014825; AAN37172.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 1163 AA; 139588 MW; 6B37EB4621D0DAB1 CRC64;

Alignment Scores:
Pred. No.: 8.61e-05 Length: 1163
Score: 172.00 Matches: 195
Percent Similarity: 32.13% Conservative: 135
Best Local Similarity: 18.99% Mismatches: 336
Query Match: 2.24% Indels: 361
DB: 5 Gaps: 54

US-09-874-162a-7 (1-4409) x Q8IKP5 (1-1163)
QY 106 TCGGGACTCCATTCCTCCCACTGCGGACCTCATCGACCATCGAGGACCAACCATC 165

```

```

Db      271 CysGlyLeuArgPheMetThrSerMetAspLysLeuAsnHisLeuGlu---AsnHisTyr 289
QY      166 GATACAGATCCACGGGTTTGTAGAAAACAAGAATTACAGCAGCAAC- 213
Db      290 LysLysSerGlnPheTyrLeuAsnAsnSerGluArgSerPheThrLysSerLysLys 309
QY      214 ---TATGTTCCTCCAGCTTACATAAA- 237
Db      310 ArgPheLeuTyrLeuAspHisIleAsnLeuProIleGluValPheValCysLysAsnTyr 329
QY      238 -----AGATTTC 243
Db      330 SerilePheGluAspPheTyrAspAsnValIleThrLysAsnMetAlaSerPheAsnPhe 349
QY      244 ATG-----ACAGATGCTGCCCGC 261
Db      350 GlnGlySerHisIleGluGluGlnIleLeuAspAsnAsnAspAsnSerTyrHis 369
QY      262 CGAGAGCAGGAGTCCCTTA-----AAGAAGAAGATTTCAGCCGAAG 300
Db      370 AsnAspAspGluLysIleTyrTyrGluGluGlyLysLysLysAspGluSerLys 389
QY      301 CTCTCGCTGACTGTCTCCAGCTCAGTCTCTCGAGGGAATGTCCCACTCCCGCAGCCAC 360
Db      390 AsnAsnAspSerLeuTyrSerTyrLeu---HisThrLysIleHisThrHisProAsnLys 408
QY      361 AGCAGTGGAAAGCCTTACTCCCGCGTGACCCCATCACCCCTCTCTCTTCATTCGCG 420
Db      409 AspLysAsn-----GlnGluSerPheLeu 416
QY      421 AGCAGCACTCGGACAGAGCCACACAGATCTATAGATTTCTTGAATCTCGAATCTCATA 480
Db      417 AsnAsn-----AsnThrLysAsnIleGln 424
QY      481 GCACCAATATTTTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCCAGAACAAAC 540
Db      425 GluLysTyrAsnValTyrAsnMetLeuSerTyrAsnAsnAsnAsnAsnAsnAsp 444
QY      541 ATCAAA-----AGAAAACATTTAAAGTTGATGATATGTTATCAAAAGTA 585
Db      445 ValAlaAsnLeuPheAsnPheLysLysThrTyrAsnAsnIleAsp----- 459
QY      586 GAGAAATGAAAGAGAGAGCAAGATCTCATAGTGTGTGACGTCATTTGCAGCTTACGTTT 645
Db      459 ----- 459
QY      646 ACTGGTTTCTTCCACAAAATGATAGCCATCACCACCAACTCAGAAAATGACAAAATTC 705
Db      460 ---AsnPhePheHis-----AsnLysGluThrIleHisAsnMet 471
QY      706 GTTACCTCGGAAGTCTGCTGTGGAAGTTTGCACAAAAGAAAGAGATGTAAGTTGT 765
Db      472 IleAsnProAsnMetAlaValAspLys-----AsnLysLysLysArgAsnIle- 487
QY      766 CCAATAAGGCAAGTTCCACAGGTAAAGAGCGTGTCTTGATCTCTGACCTCAATCAA 825
Db      488 -----LeuAsnAsp 490
QY      826 ACAAAACCCGGAATTTCCCGTCCCTTGCAGTTTCCAGTAAATGAATTTGAACTAGTAAC 885
Db      491 ArgLysLysGly-----SerAsnGluIleAsnPheAspTyrIleTyrGlyAsn 507
QY      886 AGCATATGTGAAGTCTTAC---TCGTTGCTATTTAGAGTGACTCGTCCAGGAAGAAGA 942
Db      508 GluAsnThrTyrAspValTyrLeuSerAsnTyrTyrThrValThrGluAspAsnSerVal 527
QY      943 GAGTTTAATGGAATGATTAATGAGAAACC---AATGAAAATATT----- 984
Db      528 LeuIleAsn---TyrIleAsnGlyProThrMetTyrArgAsnIleMetLysCysLeuGlu 546
QY      985 -----GATGTC---AATGAAGAGCTTCAGCC-----AGAAGAAAACGAAATCGT 1026

```

547	IleLysAspIleTy rAsnTy rLysPheProSerTrpIleProLysArgSerIleAsn---	565
1027	GAGGATGGGGAAGACATTTGTTGCAAAATGACAGTATTTGATATAAAACAGGCGCTTA	1086
566	-----AsnPhePheIleLysArgIleIleGluIleAspLysAsp-----	578
1087	CAGCTTTTAGATGGGGAATATGAAGTACCATGCAGGAAATGGAAGAATGTCCTCATRAGC	1146
579	-----LeuIleGly-----IleGln	583
1147	AAAGAAAGACAAACATGGGACTATTCTTGATGGGAAGAGGCTGCTCCATTCGAAACA	1206
584	LysLysThrAsnIlePheThrAsnIleLeuSerGlyArg-----ThrProSerAsnThr	601
1207	TTTTCT---CAGGGACCTACCTTCGACTTCCTTCGTTGGACGAGGAGACCAATGAT	1263
602	PheProLeuAsnSerTy rAsnIleGlnAspAspLeuAsnIleAsnAsnAsnThrAsnAsn	621
1264	AAATCTACGGCTCCTATTGTCACAAACCTCTTGCCACTAGAAAATTCAGAGAGCTGC---	1317
622	AsnLeuLysGlyLysAsnLysLysLysIleAspValIleAsnLysAsnLeuTrpAsn	641
1318	-----CATCAG---GAAAAACAAGCCTGGTTCAGTTAAACCTACTCAAACTATT	1362
642	IleCysThrThrHisGluPheTy rAsnSerLysGlyLeuCysProIleAsnValPheLeu	661
1363	GCTGTTTAAAGAATCATTCGACTACAGATCTACAACA-----AGA	1401
662	SerTy rGlnLysGluAsnThrLeuAsnHisGlnThrGluHisSerAspLysAsnAspHis	681
1402	AAAAAGGATCTCCAAAT-----GAAAAAC	1428
682	AspAspLysGluGluSerAsnAsnTy rProMetGlnAsnSerIleCysThrGlnGluSer	701
1429	CGACAAAATTAAGAAATATTTATCAGTTCTCTATPACAAATACAGGCAACAACT	1488
702	ValGluAsnAsnHisValLeuTy rSerPheLeu-----AsnIlePheAsnGlnLysTy r	719
1489	GAAGCAAGAGATGACTGCTCATTTGCCCTTGGTGTACTCTGAATCGCGCAAACTTTAT---	1545
720	PheAlaGlnThrAspIlePheLysAsnIleLeuLeuPheHisIleArgLysHisTy rPhe	739
1546	AGTTTACTCAAGATCTTAA-----CTCGCCATGACAGATTTATC	1587
740	MetIleIleAsnGluLeuLysAspAsnIleAspPheAsnGluLeuHisSerArg---	758
1588	TTCAACTATGTTTATCATCCAAAAGGTGCTAGATAGATT-----Leu	1629
759	PheGlnTy rIleTy rAsnAsnTy rSerLysThrIleSerIlePheAsnIleAsnAsnTy r	778
1630	TCTATCAATGAGTGTATGATGGCTCCTATGCAGGAAATCCTCAGGATATTCATCGCAA	1689
779	LysIleAspThrCysPhe-----LeuCysLys	787
1690	CTGTGATTGCTTTTAGTCGCAACGGACCATTTAAGAGAACCTATCACACATATTCCT	1749
788	GluAsnPheSerPheGluTy rSerTy rGluTy rAsnAspPheTy rTy rAsnValIle	807
1750	GTGTGCGCGCAAAACGAAACAAAAGCAAGCATGCTCGAATTTCTGAATCTCGAATGGG	1809
808	CysValAspLeuLys-----AsnValTy rGluAsnAspValGluAspThr	823
1810	GAAGTAGAACAGCAAGAACATATAGTAGTGGC-----	1842
824	GluGluAspAlaAsnLysAsnTy rAspIleAspMetSerIleLysArgIleAspHisMet	843
1843	-----CACAACTCGTCTGATTTTCCATAGTAGTACTGCTTACCTCTC	1884
844	CysAspGluTy rValTy rAsnAsnAsnSerTy rAspIleMetAspLysCysLeuTy rGlu	863
1885	CGTCCACAAGAAATG---GAAGTAGATAGTGAAGATGAAAGGATCCTGAATGGCTAAGA	1941
864	LysHisLysGluLeuAsnGluLeuLeuTy rGluAsnLysGlu-----MetSer	880

Qy	1942	GAATAAACCAATTACACAAATGAAGAGTGT	-----TCTGATGTGAAT	1981
		:::       ::    ::	::    ::	
Db	881	AspAspAlaIleThrGlnValGluLeuLeuAsnAsnLysMetValAsnAsnGluIleAsn	-----	900
Qy	1984	GAA-----GGAGAGAAAGAGTGAAGTCTGGAATCTGCAATCTCCATGTCATGAAGCAT	--- 2034	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	901	GluCysSerAsnHisAsnAsnIleLysLysValAsnAsnIleAsnThrTyAspAsnIle	920	
Qy	2035	-----GGGTTTATTGCTGCACAAATCAAAATGAATCAT	2064	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	921	AsnIleAsnAsnIlePheThrAspIleLysLysIleAspAsn	----- 936	
Qy	2065	GCCTGTATGCTGTTTGTAGAAAATTATGACACAGAAATTAATTAAGAAGAAATTTATGTCGA	2124	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	937	---CysPheLeu-----GlyAsnThrIleIleGluSerAsnMetAspCys	950	
Qy	2125	AACTTCATGCTTCATCTACGTCACGATG-----CAT	2154	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	951	AsnPheHisMetAsnIleAsnLysLysAsnLysLysAspHisIleMetAsnHis	970	
Qy	2155	GACTTTAATCTTATTAGCATAATGTCAAATAGATAAAGCTGTTACCAAGCTCCGCTGAATG	2214	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	971	TyrAsnAsnLeuPheLysAsnLeuCysIleProTyAspThrLeu	-----AspIle 987	
Qy	2215	CAGCAAAAATTAGAAAAGGGGAATCGCTTCCCTCGCAAAACGAGAAATAACTGAAGAA	2274	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	988	LeuHisArgIleLysLysGly-----AspIleThrGlyLys	999	
Qy	2275	CAAAATGGGACGACCAATCGATTTAGTAAATTAATCTCAAAACAGAGAAAGCTTTGGAAACA	2334	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1000	ThrAsnMetThrThrAsnAsnPheSerAsnAsnThrThrAsnAspGluGlyTyThrAsn	1019	
Qy	2335	GATAGTGTCTCAGGGGTT----- 2353		
		::: :::		
Db	1020	GluGluIle-AspAsnPheLeuLysAsnTyAspAsnIleThrAsnAspIleLysAsnAl	1039	
Qy	2354	-----CAAAACAGACCAAAAACAAAACCTCGAAAAGCTCTAAC	2394	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1039	avalPheGlnSerThrTyHisAsnArgSerProSerLysAsnLysAsnSerPheAsnAs	1059	
Qy	2395	CCATGTATGGACAAACACTGAAATTCATTTTAGGGAATTCATCCTCTAAG-----	2446	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1059	pIleIleMet-----SerTyIleTyAsnAspGluAsnLysLysLysGlnHi	1075	
Qy	2447	-----AATTATGTTTGTGTTTAAATCATATGTTCCAAACAGGCACCTGTAGATGAAGTAA	2502	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1075	sSerAsnIlePhePhePro----- 1081		
Qy	2503	ATGATTTCACACAGGATATTGTTATCAGGGTTCTACTTCCTTCATTTCAATATGCAGCATTACA	2562	
		:::       ::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1082	-----SerLysAsnTyThrTyThrAsnPheThrTyPheHisIleGlnCysPheLys	1099	
Qy	2563	TGTATATCACTTTTATGATGCATTAACATCTCG-----TACTTT-----AAGCA	2610	
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1099	sAsnTyIleGluTyr-----AsnIleLeuProTyTyThrPheLeuThrLysLe	1115	
Qy	2611	TGAAAAGCAATATTTCAA 2629		
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
Db	1115	uAsnAspLysTyThrPhelLys 1121		
		::    ::    ::    ::    ::    ::    ::	::    ::    ::	
RESULT 15				
Q810T3	ID	Q810T3	PRELIMINARY; PRT; 1396 AA.	
AC	Q810T3;			
DT	01-JUN-2003 (T-EMBLrel. 24, Created)			
DT	01-JUN-2003 (T-EMBLrel. 24, Last sequence update)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)			
DE	Cyclin B3.			
GN	CCNB3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae Mus.			

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Testis;  
 RA Lozano J.C., Schatt P., Picard A.;  
 RT Cloning of two mRNA coding for two isoforms of mouse cyclin B3.";  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ555464; CAD88194.1; -;  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:000074; P:regulation of cell cycle; IEA.  
 DR InterPro; IPR006670; Cyclin.  
 DR InterPro; IPR004367; Cyclin\_Cterm.  
 DR InterPro; IPR006671; Cyclin\_N.  
 DR Pfam; PF00134; cyclin; 1.  
 DR Pfam; PF02984; cyclin C; 1.  
 DR SMART; SM00385; CYCLIN; 2.  
 KW Cyclin.  
 SQ SEQUENCE 1396 AA; 158968 MW; 878PB226546F7FC0 CRC64;  
  
 Alignment Scores:  
 Pred. No.: 0.000117 Length: 1396  
 Score: 170.50 Matches: 185  
 Percent Similarity: 31.51% Conservative: 109  
 Best Local Similarity: 19.83% Mismatches: 344  
 Query Match: 2.22% Indels: 295  
 DB: 11 Gaps: 37  
  
 US-09-874-162A-7 (1-4409) x Q810T3 (1-1396)  
 QY 106 TGGGACTCCACTTCCACC-----CTGGCCGACTCATCGAGCAC 147  
 DB 160 CysAlaPheHisGlnGluThrLeuLeuMetGluLysProLeuLeuValGluThr 179  
 QY 148 ATCGAGGACACACATGATACAGATCCACGGTTTGTAGAAAAACAAGATTA---CAG 204  
 DB 180 GluAspTyrAsnGluPheAspThrGluLeuMetThrSerLysLysAspLysProGlu 199  
 QY 205 CAGCAACCTATGTTGCCCTGAGTTACATAAATAGATTATCATCAGATGCTGCCGCCGA 264  
 DB 200 AspProThrIleIle-----Glu 205  
 QY 265 GAGCAGGAGTCCCTAAAGAAGAGATTGACCGAAGCTCTCGCTGACTCTGTCC----- 318  
 DB 206 GluMetThrAspLeuLysLysSerValIleArgLysValThrLeuThrSerProLeu 225  
 QY 319 -----AGCTCA 324  
 DB 226 TrpLeuLysAsnLysHisValValGlnGluLysProValIleGlnLysSerSer 245  
 QY 325 GTGTCTCGA-----GGGAATGTCTCCACTCCGCCAGCCACAGAGTGGAGC 372  
 DB 246 PheLysArgIleSerLeuValSerAsnValValThrThrLysGlu----- 260  
 QY 373 CTATCTCCCGTACCCACCCATACCCCTCTCTTATTCCTCCGAGCAGCACTCCG 432  
 DB 261 ---LysProProValLysLysPro-----HisPheArgLysLysLysPro 274  
 QY 433 ACAGAGCCACACAGATCTATAGATTCTTCGAATCGGAATCTCATAGCACCATATTT 492  
 DB 275 Thr-----ThrGluMetLysSerLeuLeuGlnGluProSerLeu----- 287  
 QY 493 TTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCCAGAAACATCAAAAGGAAA 552  
 DB 288 -----GluGluLysTyrAsn 292  
 QY 553 ACATTAAAGTTGATGATGATTATCAMAAGTAGAGAAAANTGAAGGAGAGCAGAATCT 612  
 DB 293 ThrGlnGluAspAlaSerIleLeuLysLysProGlnValLeuGlnGluAsnThrAsn 312  
 QY 613 CATAGCTCTGACTCAATTGTCAGCTTACGTTTACTGTTTCTTCCACAAAATGAT--- 669  
 DB 313 LysAspAlaThrLeuThrGluProValThrPheLysGlyLysHisSerAlaAsnGluAla 332

670 -----AAGCCATCAACCAAACTCAGAAAAT-----GAACAAAATTCGTGT 708  
 DB 333 ThrHisThrLysLysProSerSerSerLysAsnAsnProAspProGlnGlnLysGlyThr 352  
 QY 709 ACCCTGGAAGTCTCTGTGTG----- 729  
 DB 353 AsnLeuArgProLeuArgValHisProValThrTyrGluAsnGluProMetSerSerLys 372  
 QY 730 AAGTTTGGCCACAAAAAAGA-----AAGGATGTA 759  
 DB 373 LysSerThrThrLysLysAspSerHisPheHisGlyProSerValLeuProAspLys 392  
 QY 760 AGTTGTCCAATAAGCAAGTCCACAGAGTAAAGCAGGTCTTTGATTCCTGACCTC 819  
 DB 393 HisSerProGlnMetGluValSerThrValLysSerLeuAlaLeuProAsnProThr 412  
 QY 820 AATCAAAACAAACCCGAAATTTCCCGTCCCTTCAGTTTCCAGTATGATGATTTGAACCT 879  
 DB 413 ThrGluGluLysMetLeuHisPheProValAlaThrValLeuGluLysGlnHisAsnMet 432  
 QY 880 AGTAACAGCCATATGTTGAAGTCTTACTCGTGTCTATTAGAGTACTCGT---CCAGGA 936  
 DB 433 GlyGluAlaProCysLysLeuLysLysProSerProLeuArgLysGlnGlnLeuProLys 452  
 QY 937 AGAAGAGAGTTTAAATGGAATGATTATGAGAGAACCAACATGAAATAT--- 984  
 DB 453 ArgArgArgPhe-----PheSerAsnSerAlaValGlnGluThrValIleArgLysPro 470  
 QY 985 -----GATGTCAATGAAGAGCTTCCA---GCCAGA 1011  
 DB 471 LeuPhePheLysMetSerThrThrGluLysAspProSerGlnTrpProSerAlaLeu 490  
 QY 1012 AGAAACGAAATCGTGAAGTGGGAAAGACATTTGTCACAAATGACAGATTTTGAT 1071  
 DB 491 ProLysLysHisIleSerProGlyGluLeuSerLysGlnLysLysGlnHisValSerPro 510  
 QY 1072 AAAAAACGCGCTTACAGCTTTTAGTGGGGAATATCAAGTCCCTGACAGGAAATCGAA 1131  
 DB 511 LysHisAsn-----MetGluGluAspSer 518  
 QY 1132 GAATGTCCAATAAGCAAGAAAAAGACAACTGGGAGACTATTCTTGTGGAAAGAGCTG 1191  
 DB 519 GlnCys-----TipLeuAspSerAlaPheLysLysGlnLeu 530  
 QY 1192 CTCTCCATCGAAACATTTCTCAGGACCTACGTTGCAGTTCTACTCTCTCTG--- 1245  
 DB 531 SerArgGluGluProAlaSerThrHisThrProLeuLysLeuGluMetGlnGlnAlaIle 550  
 QY 1246 ACAGGAGAGACCAATGATAAATCTACGGCTCTATTGCCCCAACTCTTGCACCTAGAAAT 1305  
 DB 551 ThrLysGluThrGlyPheHisLeuArgAsnProLeuValLeuProThrValThrSerGlu 570  
 QY 1306 TCAGAGAGTCTCCATCAGGAAAAACAAGCTGTTTCAAGTTAAACCTACTCAAACTATTGCT 1365  
 DB 571 AlaLysSerLeuThrLysGlu---ProProSerPheArgGluGlnAsnThrSerLeu 588  
 QY 1366 GTTAAAGAAATCATGTACAGATCTACAAACAGAAAGAAAGAGGATAC--- 1416  
 DB 589 LeuLysArgLysSerThrThrHisThrIleThrLeuGlnGlnAlaGlnSerGluTrpGln 608  
 QY 1416 ----- 1416  
 DB 609 GluMetThrAspGluAspArgAsnLeuPheSerIleLysProGlySerHisArgLysGlu 628  
 QY 1417 -----CCAAATGAAAAAC-----CGACAA 1434  
 DB 629 ProIleProGluPheLeuGlnAsnProLeuProAsnGluAsnCysLeuIleSerGln 648  
 QY 1435 AAATTAAAGATATTTTATCAGTTCTCTATACAAACATACAGGCAACAACTGAAGCA 1494  
 DB 649 LysLeuSerHisSerMetProPheAlaSerGlnLysThrThrSerGlnGluArgAlaHis 668  
 QY 1495 AGA-----GATGACCTGCATTGC 1512

Job time : 625 secs

```
Db      669 ArgLysGluSerValAlaSerAsnAspLysAsnPhePheSerGlnAspLeuPheSer 688
QY      CTTGGTGTACTGAACTGCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGC 1572
Db      689 ProPheSerSerAlaAspGluAspThrLeu----- 698
QY      1573 CATACAGATTATCTTCAACTATGTTTATCATCCAAAGGTGCTAGATAGATGTTTCT 1632
Db      699 -----LysPheHisLysSerLeuAspPheGlnGluValAlaAspArgLysAsnAspSer 716
QY      1633 ATCAATGAGTGTATGATGCTCTATGCAGGAATCTCAGGATATTCATCGCCAACT 1692
Db      717 HisLysLysMetPheAspSerGlnAspSerValSerGluGluGlu----- 731
QY      1693 GGATTGCTTTTAGTCGCAACGGACAGTTAAGAGAACACCTATCACATATTCCTTGTG 1752
Db      732 -----SerPheLeuArgLys-----LeuPhe 738
QY      1753 TCCAGGCCAAAA-----CGAACAAAGCAAGCATG 1782
Db      739 CysLysAspArgCysSerSerThrGluGluLeuSerGlnGluArgThrValAlaLeuGlu 758
QY      1783 TCTGAATTTCTT-----GAATCTGAAGATGGGGAAGTAGAACACAGCAA 1824
Db      759 GlnGluPheLeuLeuLeuLeuLeuAsnThrSerSerAspValAspGluPro 778
QY      1825 AGACATATAGTAGGCGCAACGCTCTGTATTTCCATAGTATGATGCTGCTACCTCTC 1884
Db      779 LeuSerHisGlnSerProHis-----IleGlnAsnHisSerAspThrThrLysGluAla 796
QY      1885 -----CGTCCACAGAAATGAGAA-----GTA 1905
Db      797 LeuGluAlaSerGluAlaLeuGluAlaProGluAlaLeuGluThrLeuGluAlaLeuVal 816
QY      1906 GATAGTGAAGATGAAAGATCCTGAAATGCGTAAAGAGAAAA----- 1947
Db      817 AlaSerGluAspLeuGluGluProLeuAsnIleLeuGluLeuLeuSerThrGluAsnMet 836
QY      1948 -----ACCATTAACAATGAAGAGTTTCTGATGTTAATGAGGA 1989
Db      837 ValAlaLeuMetLysMetLeuValThrGluAspGluSerThrLysAspSerPheSerGly 856
QY      1990 GAGAAAGAGTGAAGAACTCGAATCTCGATCATGAAGCATGGGTTTATGCTGAC 2049
Db      857 AsnTyrThrAlaAlaArg-----GluAlaHisAlaGluLysSerLeuSerLeuGluGlu 874
QY      2050 AATCAATGAATCATGCTGCTGATGCTGTTGTAGAAATATTGACAGAAATAATTAAG 2109
Db      875 ThrSerIleAsnGluAlaAlaThrLeuLysGluSerLeuSerGlnGluLysHisArg 894
QY      2110 AAGAAATTTATGCGAAACTTCATGCTTCATAGTCAGCATGACATTTAATCTTTATT 2169
Db      895 AlaGluLeuValThrValLeuLysGluLeuValLeuMetLysAsnProSerLeuLys 914
QY      2170 AGCATATG-----TCATAGATAAGCTGTACCAAGCTCCGTGAA 2211
Db      915 ArgValAlaLeuAlaPheGlnGluAsnProSerAsnAsnValGluThrLeuLeuArgGlu 934
QY      2212 ATGCAGCAAAATAGAAAGGGGAATCTGCTTCCCTCGAAACGAAAGAAATAACTGAA 2271
Db      935 ValLeuAlaLeuValGluAsn-----SerThrAlaAspGluSerThrLeuGln 950
QY      2272 GAACAAATGGCAGACAGCAAT-----GGATTAGTGAAT 2307
Db      951 GluLysProSerThrLysThrAspValThrProLysGluLeuLeuAlaLeuGluGluAsn 970
QY      2308 AACTCAAAAGAGAAAGCTTTGAAACAGATAGTGTCTCA 2346
Db      971 SerSerAsnLysLysAlaAsnProMetAspSerLeuSer 983
```